PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/80, 1/15 // (C12N 1/15, C12R 1:665, 1:685, 1:77, 1:885)

A1

(11) International Publication Number: WO 99/32641

(43) International Publication Date: 1 July 1999 (01.07.99)

(21) International Application Number: PCT/EP98/06519

(22) International Filing Date: 6 October 1998 (06.10.98)

(30) Priority Data:
97204062.0 22 December 1997 (22.12.97) EP

(71) Applicant *(for all designated States except AU BB CA CY GB GD GH IE KE LK LS MN MW NZ SD SG SZ TT UG)*: UNILEVER N.V. [NL/NL]; Weena 455, NL–3013 AL Rotterdam (NL).

(71) Applicant *(for AU BB CA CY GB GD GH IE KE LK LS MN MW NZ SD SG SZ TT UG only)*: UNILEVER PLC [GB/GB]; Unilever House, Blackfriars, London EC4P 4BQ (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: DE GROOT, Marcellus, Johannes, A. [NL/NL]; Unilever Research Vlaardingen, Olivier van Noortlaan 120, NL–3130 AC Vlaardingen (NL). BEIJERSBERGEN, Alida, Godelieve, M. [NL/NL]; Unilever Research Vlaardingen, Olivier van Noortlaan 120, NL–3130 AC Vlaardingen (NL). MUSTERS, Wouter [NL/NL]; Unilever Research Vlaardingen, Olivier van Noortlaan 120, NL–3130 AC Vlaardingen (NL).

(74) Agent: VAN DER TOORREN, Johannes; Unilever N.V., Patent Dept., P.O. Box 137, NL–3130 AC Vlaardingen (NL).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: A PROCESS FOR SITE-DIRECTED INTEGRATION OF MULTIPLE COPIES OF A GENE IN A MOULD

(57) Abstract

A process for site-directed integration of multiple copies of a gene in a mould is provided, which comprises transforming a mould cell containing in its chromosomal DNA a restriction site for a rare-cutting endonuclease, e.g., I-*Sce*I, preferably introduced at a desired locus, e.g., within a selectable marker gene or in the neighbourhood thereof, with a piece of DNA comprising multiple copies of at least one expressible gene comprising at least one structural gene encoding a desired protein, surrounded by two DNA fragments homologous to part of the DNA upstream and downstream, and in the neighbourhood, of said restriction site, while during the transformation of the mould the presence of the rare-cutting endonuclease is provided, followed by selecting or screening for a mould cell in which the multiple gene copies of said expressible gene are inserted into the chromosomal DNA of the mould. The piece of DNA can comprise a third DNA fragment completing any disrupted or partially deleted selectable marker gene in the chromosomal DNA. Preferably the mould belongs to the genus *Aspergillus*, especially to the species *Aspergillus awamori*. Also provided are a transformed mould obtainable by a process according to the invention, a process for culturing such transformed mould, and a process for producing and optionally secreteing a desired protein by culturing such transformed mould under conditions whereby the structural gene encoding said desired protein is expressed, and optionally isolating or concentrating the desired protein.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Title: **A process for site-directed integration of multiple copies of a gene in a mould**

The invention relates to a process for site-directed integration of multiple copies of a gene in a mould, to a transformed mould obtainable by such process, to a process for culturing such transformed mould, and to a process for producing and optionally secreting a desired protein by culturing such transformed mould. In particular, the invention provides a process for preparing a protein by a mould transformed by multicopy integration of at least one expressible gene comprising a structural gene encoding a desired protein into the genome of a mould, especially of moulds belonging to the genus *Aspergillus*.

In this specification the expression "**expressible gene**" means a structural gene encoding a protein, either homologous or heterologous to the host organism, in combination with DNA sequences for proper transcription and translation of the structural gene, and optionally with secretion signal DNA sequences, which DNA sequences should be functional in the host mould. Further, in this specification the expressions "**mould**" and "**filamentous fungus**" are considered as synonyms.

Background of the invention and prior art

1. Filamentous fungi and especially species such as *Aspergillus awamori, Aspergillus niger, Trichoderma reesei* and *Fusarium graminearum* have shown to be attractive hosts for large scale production of homologous and heterologous proteins. They have the capacity to secrete substantial amounts of protein into the medium, large scale fermentation is generally well established and most of them they have a GRAS (Generally Recognized As Safe) status, which makes it possible to use these species in the food and food-processing industry. Moreover, the mould *Fusarium graminearum* A 3/5, the Quorn^{R} myco-protein fungus, has also been used as a commercial human food source in the UK for

over 10 years (Royer *et al.*; Bio/Technology 13 (1995) 1479-1483).

The production of fungal proteins, of either homologous or heterologous origin, by filamentous fungi is usually very efficient and production levels of grams per litre were reached. However, compared to this the production levels of heterologous proteins of mammalian, bacterial or plant origin in moulds are relatively low. In order to improve the production of both homologous and heterologous proteins several strategies have been developed. The basic strategy that is commonly applied to achieve higher protein production in moulds is the introduction of multiple copies of the gene encoding the desired protein.

2. Whereas moulds have been successfully used for the production of enzymes, antibody fragments and peptides at laboratory and commercial scale (xylanase, pectinase, etc), the acceptance of products from these genetically modified organisms (GMO) in the market has experienced some unexpected difficulties in the past few years.

(a) In general there is a growing concern about the use of antibiotic resistance genes in genetically modified organisms. The main reason for this concern is the possibility that such a gene might be transferred into and expressed in gut micro-organisms, which would thereby become antibiotic resistant ("Report on the use of antibiotic resistance markers in genetically modified food organisms" published by the Advisory Committee on Novel Foods and Processes, Ministry of Agriculture, Fisheries and Food, England, 1994).

(b) Further, the presence of other foreign DNA such as remnants of vector DNA used in cloning is also undesired.

(c) Another concern is the fact that in general the genetically modified strains contain randomly integrated genetic material. In the perception of some consumer organisations this would constitute an unpredictable safety risk, and could mean a barrier to the acceptation of derived products.

3. Therefore, the recombinant mould should ideally contain **multiple** copies of the gene encoding the desired protein integrated at only a **predetermined** locus in the genome and **no other foreign DNA** should be present in order to produce proteins in moulds in both an economically attractive manner and in a way that deals with the concerns about genetically modified organisms as described above.

The generation of mould strains that meet these criteria has not been reported in literature.

The commonly applied system for integration of single or multiple copies of a gene into the genome of moulds, e.g. *Aspergillus*, *Trichoderma* and *Fusarium graminearum* makes use of plasmids which in addition to the gene encoding a desired protein contain bacterial marker genes encoding resistance to antibiotics (e.g. Ampicillin) and other vector sequences. Therefore, genetically modified moulds will usually contain antibiotic resistance genes and other vector DNA.

Whereas, targeted integrations of a **single** gene copy have been described regularly (e.g. Timberlake, "Gene Cloning and Analysis" (Chapter 3) in the book "More Gene Manipulations in Fungi" (1991) 51-85, edited by Bennett and Lasure; Gouka *et al*. Applied and Environmental Microbiology 62 (1996) 1951-1957) it has been proven to be very difficult to obtain mould strains that contain **multiple** gene copies integrated at a predetermined locus in the genome. Gouka *et al*. (Curr. Genet. 27 (1995) 536-540) reported the selection of targeted multi-copy integrations at the *pyrG* locus in *A. awamori*, but the recombinant strains were obtained from transformations in which DNA was used containing vector sequences and no information was presented on the number of gene copies that were integrated at the *pyrG* locus. For *Aspergillus nidulans* a similar observation on targeted tandem integration at the *argB* locus was published (Van den Hondel and Punt, "Gene

transfer systems and vector development" (Chapter 1) in the book "Applied Molecular Genetics" (1991) 1-28, edited by Peberdy *et al.*).

Several other publications indicate that **site-directed integration of multiple gene copies could not be obtained**, although it was desired for scientific or commercial purposes, (Kubicek-Pranz *et al.* J. of Biotech. 20 (1991) 83-94; Van den Hondel *et al.* Antonie van Leeuwenhoek 61 (1992) 153-160; Verdoes *et al.* Transgenic Research 2 (1993) 84-92; Archer *et al.* Antonie van Leeuwenhoek 65 (1994) 245-250; Van Gemeren *et al.* Applied Microbiology and Biotechnology 45 (1996) 755-763; Van Gemeren, "Expression and secretion of defined cutinase variants by *Aspergillus awamori*" (Chapter 5) Thesis University of Utrecht (1997) ISBN 90-393-1229-X).

4. Previously, two processes have been described in literature that, in principle, might allow the generation of mould strains that contain multiple copies of a gene that are integrated at a predetermined locus in the genome without the presence of other foreign DNA.

The **first** process describes the preparation of a protein by a fungus transformed by site-directed multicopy integration of an expression vector in the ribosomal DNA locus of the fungal genome as described in International PCT patent application WO-A-91/00920; Unilever, published 24 January 1991. Although the Examples were carried out with yeasts, it was envisaged that such process is also applicable to moulds. Thus such process could make it possible to construct a mould strain in which multiple copies of a gene are integrated at a predetermined locus of the genome, without the presence of other foreign DNA.

However, transformation of moulds follows a somewhat different pattern than the transformation of yeasts. Whereas in the yeast *Saccharomyces cerevisiae* transforming DNA is integrated into the genome of the cell via

homologous recombination at the corresponding homologous site, in filamentous fungi such as the mould *Aspergillus awamori* DNA integrates mainly via illegitimate recombination at random sites in the genome (Finkelstein, "Transformation" Chapter 6 in the book "Biotechnology of Filamentous Fungi" (1992) 113-156, edited by Finkelstein and Ball). For instance, for the mould *A. awamori* Gouka *et al.* (Curr. Genet. 17 (1995) 536-540) performed an analysis on a large number of transformants and showed that DNA integrated via homologous recombination in approximately 10 % of the transformants, whereas the remaining 90 % integrated randomly. This means that transformants have to be screened for site-directed integration events. Therefore, a process for transformation of moulds as described in WO-A-91/00920 would require lengthy screening procedures because DNA that is introduced into the mould cell can also integrate randomly and not only via homologous recombination at the predetermined site.

The **second** process describes the site-directed integration of a **single** gene copy whereby any other heterologous DNA used for cloning and any heterologous mould selection marker are removed, as described in European patent application EP-A1-0 635 574; Gist-brocades N.V., published 25 January 1995.
If this second process would be used for preparing a transformed mould containing multiple gene copies, the process is very cumbersome, because the whole process need be repeated for each subsequent copy that needs to be introduced.
Although the repetition of the second process for obtaining multicopies is mentioned as simple statements in the specification (see e.g. page 3, lines 22-23, page 6, lines 21-25, page 7, lines 29-30, and page 8, lines 9-11, and 15-16), it was not shown in the Examples that it really works. In fact the statement "sequential application of the same technology" mentioned on page 8, lines 9-11 confirms the laborious character of this method for introducing multiple

gene copies at predetermined loci, covering both a single site and multiple sites.

A further disadvantage of the method described is the risk that the earlier introduced desired foreign DNA is removed during a subsequent repetition of the process.

In summary, items 1-4 above show that there exists a need in the field of mould biotechnology to construct mould strains containing multiple copies of a gene encoding a desired protein that are integrated at a predetermined locus in the genome and that are free of bacterial antibiotic resistance genes or of other foreign DNA such as remnants of vector DNA used in cloning. Ideally, the recombinant microorganism should only contain the heterologous gene encoding the desired protein.

Summary of the invention

The invention is applicable in the field of mould biotechnology and provides a new and more advanced process for site-directed integration of multiple copies of a gene in a mould without leaving any undesired DNA, i.e. without leaving in the transformed mould the selection marker used for selection of transformants or other DNA used for cloning. The invention is based on the specific introduction of a double-strand break at the chromosomal target in the mould cell which significantly enhances site-directed integration at that locus. Repair of the break with a repair DNA homologous to the regions flanking the break and including multiple copies of at least one gene encoding at least one desired protein will lead to simultaneously integration of those multiple copies at the locus of the break.

The present invention provides a process for transforming a mould, in which

(1) multiple copies of a desired gene are integrated in the chromosome of said mould,

(2) the integration in the mould genome is site-directed via homologous recombination in contrast to the usual random integration of moulds,

(3) such site-directed integration event is selected preferentially over any possible random integration event, e.g. by selecting for the restoration of a defective marker gene,

(4) remaining foreign DNA sequences, e.g. antibiotic resistance genes and DNA originating from other organisms, **can** be avoided, and

(5) a rare-cutting endonuclease, e.g. I-*Sce*I, is used to introduce a double-strand break in the chromosomal DNA of the mould,

Although the emphasis is given to the use of I-*Sce*I as a rare-cutting endonuclease, it is envisaged that also other rare-cutting endonucleases can be used, including HO Endonuclease and VDE, the latter also being known as PI-*Sce*I.

Brief description of the drawings

Figure 1. shows the construction of the plasmids pUR5710, pUR5711 and pUR5712, in which
amp = ampicillin resistance gene, and
pyrG = *pyrG* gene from *A. awamori*, '*pyrG* or *pyrG*' indicates that the gene is truncated at the 5' or 3' end, respectively.

Figure 2. shows the construction of the plasmids pUR5713 (Figure 2A) and pUR5714 (Figure 2B), in which
pBS-SK = pBluescript^{R}-SK.

Figure 3. shows the construction of the plasmids pUR5716 and pUR5718, in which cos = *cos* site.

Figure 4. shows the construction of the plasmid pUR5729, in which
PexlA = Promoter sequences of the *A. awamori* 1,4-β-endoxylanase A gene,

cut = coding region of the *F. solani pisi* cutinase gene (synthetic copy of cDNA), and
TexlA = Terminator sequences of the *A. awamori* 1,4-β-endoxylanase A gene.

Figure 5. shows the construction of the cosmids pUR5722 and pUR5725.

Figure 6. shows the construction of the plasmids pUR5736 and pUR5737, in which
Pgpd = Promoter sequences of the *A. nidulans gpd* gene,
hph = coding region of the hygromycin phosphotransferase gene from *E. coli*, and
TtrpC = Terminator sequences from the *A. nidulans trpC* gene.

Figure 7. shows the construction of plasmid pUR5724, in which I-SceI = synthetic gene encoding the *Saccharomyces cerevisiae* I-*Sce*I endonuclease.

Figure 8. Experimental design of the process for site-directed integration of multiple copies of a gene in the mould *A. awamori* using the I-*Sce*I endonuclease. The wild-type *pyrG* gene is depicted in the upper part of the figure. The coding region of the gene is indicated by the light grey shaded box. Below this, the target locus containing the
I-*Sce*I restriction site as present in the *A. awamori* strain AWCSCE is shown. Between the non-functional 5' part of the *pyrG* gene and 3' flanking sequences of the chromosomal *pyrG* locus an I-*Sce*I site is present. The fragment that is introduced into the strain AWCSCE contains a non-functional 3' part of the *pyrG* gene that is partially homologous to the mutated *pyrG* gene at the chromosomal target locus, one or multiple gene copies (indicated by the dark grey shaded boxes 1,2 and n) comprising at least one structural gene encoding at least one desired protein and an additional sequence from the *pyrG* locus that is homologous to

sequences present immediately downstream of the I-*Sce*I site at the target locus. Simultaneously, the I-*Sce*I endonuclease or an expression plasmid containing the I-*Sce*I gene is co-introduced into the cell. After homologous recombination induced by a double-strand break at the I-*Sce*I site an intact *pyrG* gene is restored and the multiple gene copies are simultaneously integrated at the *pyrG* locus, which is illustrated in the lower part of the figure.

Figure 9. shows the autoradiograph of the Southern blot of *A. awamori* genomic DNA probed with an 18 bp end-labelled oligonucleotide representing the I-*Sce*I restriction site. The genomic DNA was digested with *Sau*3A. M represents the 1 kb DNA marker (BRL), lanes 1, 2 and 3 contain samples of the plasmid pSCM522 digested with *Hinf*I (control DNA substrate containing the I-*Sce*I restriction site, supplied with the I-*Sce*I endonuclease from Boehringer Mannheim, cat. no. 1497235) in concentrations that correspond to 200, 20 and a single copy of the I-*Sce*I restriction site(s) in the genome of *A. awamori*, respectively. Lanes 4 and 5 contain *Sau*3A digested genomic *A. awamori* DNA (7.5 µg). Lanes 6, 7 and 8 contain samples of plasmid pUR5712 in concentrations that correspond to a single copy, 20 and 200 copies of the I-*Sce*I restriction site(s) in the genome of *A. awamori*.

Figure 10. shows the autoradiograph of the Southern blot of the two *A. awamori* mutant *pyrG* strains. The genomic DNA was digested with *Bgl*II and I-*Sce*I and probed with a 2.4 kb *Bam*HI x *Hin*dIII fragment from plasmid pAW4.1 containing the *A. awamori pyrG* gene. M represents the 1 kb DNA marker (BRL), Lane 1 and 2 contain genomic DNA of the *A. awamori* mutant *pyrG* strains. Lane 3 contains genomic DNA from the non-transformed wild-type *A. awamori* strain.

Figure 11. Southern analysis of the recombinant strains obtained in the transformation of *A. awamori* strain AWCSCE (see Example 4). The genomic DNA was digested with *Sal*I or *Bgl*II and probed with a number of different probes (see

Example 5). The two probes that are used in the Southern blots shown in Figure 12 are indicated;

pyrG = a 2.4 kb *Bam*HI x *Hin*dIII fragment from pAW4.1 containing the *A. awamori pyrG* gene.

TexlA = 0.5 kb *Afl*II x *Sac*I fragment from pUR5729 containing the *exlA* terminator.

The wild-type *pyrG* gene is depicted in the upper part of the figure. The *Sal*I digestion will give a 3.3 kb and a 3.8 kb fragment which fragments will hybridize with the *pyrG* probe. The *Bgl*II digestion will give a 9.0 kb and a 2.7 kb fragment, which fragments do not hybridize with the *TexlA* probe.

Below this, the target locus containing the I-*Sce*I restriction site as present in the *A. awamori* strain AWCSCE is shown. The *Sal*I digestion will give a 3.3 kb and a 3.0 kb fragment of which only the 3.3 kb fragment will hybridize with the *pyrG* probe. The *Bgl*II digestion will give a 9.0 kb and a 2.7 kb fragment, which fragments do not hybridize with the *TexlA* probe.

The lower part of the figure shows a restored *pyrG* gene containing none or multiple gene copies of the cutinase gene. The *Sal*I digestion will give a 3.3 fragment and in recombinants obtained with pUR5718 a 3.7 kb fragment, in recombinants obtained with pUR5722 a 9.6 kb fragment and in recombinants obtained with pUR5725 a 17.1 kb fragment. All these fragments will hybridize with the *pyrG* probe. The *Bgl*II digestion will give a 9.0 kb fragment and in recombinants obtained with pUR5718 a 2.6 kb fragment, in recombinants obtained with pUR5722 or pUR5725 a 1.7 kb or a 1.9 kb fragment (depends on the orientation of the cutinase gene compared to the *pyrG* gene) and a 1.5 kb fragment, which is derived from the tandem repeat of the cutinase gene. Only the fragments containing the *exlA* terminator present in the cutinase expression cassette will hybridize with the *TexlA* probe.

The bottom part of the figure show the position of the *pyrG* and *TexlA* probes.

Figure 12. shows the autoradiograph of the Southern blot analysis of the recombination events. M represents the 1 kb DNA marker (BRL),
Lane 1: recombinant AWC-pUR5718#S1;
Lane 2: recombinant AWC-pUR5718#S2;
Lane 3: recombinant AWC-pUR5718#1;
Lane 4: non-transformed wild-type *A. awamori* strain;
Lane 5: *A. awamori* mutant *pyrG* strain AWCSCE;
Lane 6: recombinant AWC-pUR5725#1;
Lane 7: recombinant AWC-pUR5725#2;
Lane 8: recombinant AWC-pUR5722#A1;
Lane 9: recombinant AWC-pUR5722#A2;
Lane 10: recombinant AWC-pUR5722#B1;
Lane 11: recombinant AWC-pUR5722#B2;
Lane 12: recombinant AWC-pUR5722#B3.
A. The genomic DNA is digested with *Sal*I and probed with the *pyrG* probe (see Figure 11).
B. The genomic DNA is digested with *Bgl*II and probed with the *TexlA* probe.

Detailed description of the invention

The invention provides a process for site-directed integration of multiple copies of a gene in a mould, which comprises

(i) providing a mould cell containing in its chromosomal DNA a restriction site for a rare-cutting endonuclease,

(ii) transforming such mould cell with a piece of DNA comprising in the 5' to 3' direction in the following order

(a) a first DNA fragment homologous to part of the DNA upstream and in the neighbourhood of the restriction site for the rare-cutting endonuclease present in the chromosomal DNA of the mould

(b) multiple copies of at least one expressible gene comprising a structural gene encoding a desired protein,

(c) a second DNA fragment homologous to part of the DNA downstream and in the neighbourhood of the restriction site for the rare-cutting endonuclease present in the chromosomal DNA of the mould,

while during the transformation of the mould the presence of the rare-cutting endonuclease is provided,

(iii) selecting or screening for a mould cell in which the multiple gene copies of said expressible gene are inserted into the chromosomal DNA of the mould.

For site-directed integration it is desirable to use a restriction endonuclease forming a double-strand break at the target site, that does not form breaks at other loci in the chromosome, thus a **rare-cutting endonuclease**, an example of which is the I-*Sce*I endonuclease from *Saccharomyces cerevisiae*. The nucleotide sequence encoding this enzyme and some uses of that sequence are described in International PCT patent application WO 96/14408; Institut Pasteur, published 17 May 1996. But also other known rare-cutting endonucleases can be used in a process according to the invention, e.g. VDE also known as PI-*Sce*I (see M. Jasin; Trends In Genetics (TIG) 12 (No. 6, June 1996) 224-228; Genetic manipulation of genomes with rare-cutting endonucleases and M. Brenneman *et al.*; Proc. Natl. Acad. Sci. USA 93 (1996) 3608-3612; Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site-specific endonucleases) and HO Endonuclease (see M. Chiurazzi; The Plant Cell 8 (Nov. 1996) 2057-2066; Enhancement of Somatic Intrachromosomal Homologous Recombination in *Arabidopsis* by the HO Endonuclease). Moreover, if a specific mould genome is practically free from restriction sites for a more familiar restriction endonuclease, such endonuclease can be used as

well and can be considered a rare-cutting endonuclease for that specific mould genome.

The structural gene encoding the desired protein which gene forms part of the expressible gene can be homologous or heterologous to the mould.

In some cases the restriction site for the rare-cutting endonuclease occurs naturally in the chromosal DNA of the mould. But if the mould does not contain a restriction site for such rare-cutting endonuclease, the restriction site for the rare-cutting endonuclease can be introduced at a desired locus.

The selection of the transformed mould can be carried out by using a selectable marker. Preferentially, such selectable marker is a characteristic of a naturally-occurring, wild-type mould strain, while the mould strain to be transformed is a mutant strain thereof, deficient in said selectable marker, e.g. the orotidine-5'-phosphate decarboxylase gene (*pyrG* gene) which is present in wild-type *Aspergillus awamori*. But also other loci containing auxotrophic markers including *trpC*, *argB*, and *niaD* genes can be used, whereas other possible selectable markers include genes producing an easily assayable product. Sometimes the DNA introduced into the mould can be used as the selectable marker. For example, when the introduced DNA is expressed, it can result in a product not produced in the non-transformed mould, but which is more or less easily assayable. Or the presence or absence of the DNA can be determined by applying PCR techniques.

Preferably, the desired locus is within a selectable marker gene or in the neighbourhood thereof. In order to complement any disrupted or (partially) deleted gene the piece of DNA used for transforming the mould cell can comprise a third DNA fragment that completes any disrupted or (partially) deleted selectable marker gene in the chromosomal DNA. This would allow direct selection of strains containing the desired targeted integration of multiple copies of the gene.

The part of the DNA **up-stream** of the restriction site for the rare-cutting endonuclease present in the chromosomal DNA of the mould, to which the **first** DNA fragment is homologous, can be part of a selectable marker gene. Alternatively, the part of the DNA **down-stream** of the restriction site for the rare-cutting endonuclease present in the chromosomal DNA of the mould, to which the **second** DNA fragment is homologous, is part of a selectable marker gene. Or both the up-stream and down-stream parts can be part of the same selectable marker gene.

If two or more restriction sites for the rare-cutting endonuclease are present, the number of integrated gene copies can be increased or several different genes can be introduced at different loci, or both.
Preferably, the expressible gene comprises (1) a promoter operable in said mould, (2) optionally a DNA fragment encoding a secretion signal peptide facilitating the secretion of said desired protein from said mould, (3) a structural gene encoding said desired protein, and (4) optionally a terminator operable in said mould, whereby the promoter and the optional terminator control the expression of the structural gene. More preferably, the promoter, secretion signal and terminator are homologous to the mould to be transformed. In that case the amount of foreign DNA is kept to a minimum.

During the transformation of the mould the rare-cutting endonuclease can be provided by adding the endonuclease as such in way that is similar to Restriction Mediated Integration (REMI; Kuspa and Loomis, Proc. Natl. Acad. Sci. USA 89 (1992) 8803-8807; Redman and Rodriguez, Exp. Mycol. 18 (1994) 230-246). This is preferred when the amount of foreign DNA introduced into the mould should be as low as possible.

But for other reasons it can be convenient to form the rare-cutting endonuclease *in situ* by co-transforming the

mould with DNA encoding the rare-cutting endonuclease, which DNA can be expressed during or after the transformation of the mould. Preferably, this DNA forms part of a plasmid that does not integrate in the genome, so that after further culturing the transformed mould strain can loose the plasmid while the desired DNA is still maintained in the genome. This event can be checked by further screening to confirm the absence of the rare-cutting endonuclease-encoding DNA in the genome of the recipient strain.

Preferably the mould belongs to the fungal division of *Eumycota*, more preferably to one of the fungal sub-divisions *Ascomycotina*, *Basidiomycotina*, *Deuteromycotina*, *Mastigomycotina*, and *Zygomycotina*. It is especially preferred that the mould is selected from the genus *Aspergillus*, more particularly belonging to the species *Aspergillus awamori*. The invention also provides a transformed mould obtainable by a process according to the invention for the site-directed integration of multiple copies of a gene in a mould. Once a transformed mould according to the invention has been obtained, such transformed mould can be used in a process for further culturing.

The invention also provides a process for producing and optionally secreting a desired protein by culturing a transformed mould obtainable with a process as described above under conditions whereby the structural gene encoding said desired protein is expressed, and optionally isolating or concentrating the desired protein.

One way of introducing multiple copies of a gene is introducing several copies of the complete expression cassette as described below under the heading **Construction of multi-copy vectors**.

An alternative is by introducing several copies of the structural gene (polycistronic). After production of the

encoded polypeptide it has to be cleaved to form the single peptides, e.g. by using the enzyme KEX 2.

The invention is exemplified by the following Example preceded by a description of the Materials and Methods that were used. In this Example the following is described.

Example 1A. Experimental design of the process for site-directed integration of multiple copies of a gene in the mould *A. awamori* using the I-*Sce*I endonuclease.

Example 1B. Determination of the occurrence of a natural I-*Sce*I restriction site in the genome of *A. awamori*.

Example 1C. Construction of the *A. awamori* mutant *pyrG* strain AWCSCE which contains an I-*Sce*I restriction site at the locus of the mutated *pyrG* gene.

Example 1D. Induction of site-directed integration at the *pyrG* locus by I-*Sce*I expression.

Example 1E. Southern blot analysis of recombination events.

MATERIALS AND METHODS

Bacterial and mould strains

For standard bacterial cloning the *Escherichia coli* strain DH5α (genotype: F$^-$, *endA*1, *hsdR*17 (r_k^- m_k^+), *supE*44, *thi*-1, lambda$^-$, *recA*1, *gyrA*96, *relA*1, Δ(*argF-lacIZYA*)U169, *deoR* (phi80d(*lacz*)ΔM15); Hanahan; J. Mol. Biol. 166 (1983) 557-580) was used. For cloning multiple copies of a gene in a cosmid vector via packaging the *Escherichia coli* strain 1046 (Cami,B. and Kourilsky,P.; Nucleic Acids Research 5 (1978) 2381) was used.

The mould strain *Aspergillus awamori* #40 (a derivative of *A. awamori* CBS 115.52 also mentioned in WO 93/12237, page 9 line 13) was used to construct a *pyrG* derivative strain,

designated AWCSCE, containing an I-*Sce*I restriction site at the *pyrG* locus.

The preparation of *A. awamori* #40 (also known as *A. niger* var. *awamori* #40) was described in WO 91/19782 on page 13, lines 29-39, which read:

" The production level of the A. niger var. awamori transformants, however, can be further increased by using suitable A. niger var. awamori mutant strains, such as A. niger var. awamori #40, which produces clearly more xylanase than the wild type strain. The mutant A. niger var. awamori #40 has been obtained by mutagenesis of A. niger var. awamori spores and selection for xylanase production. In bran medium the "xylA"A. niger var. awamori #40 transformant produced 190 000 U xylanase, which is a considerable increase over the best producing A. niger var. awamori transformant. "

In this specification the following endonuclease restriction sites are used:

giving staggered ends		giving blunt ends	
*Afl*II	C↓TTAAG	*Sma*I	CCC↓GGG
*Bam*HI	G↓GATCC		
*Bgl*II	A↓GATCT		
*Eco*RI	G↓AATTC		
*Hind*III	A↓AGCTT		
*Hin*fI	G↓ANTC		
*Nde*I	CA↓TATG		
*Not*I	GC↓GGCCGC		
*Pst*I	CTGCA↓G		
*Sac*I	GAGCT↓C		
*Sal*I	G↓TCGAC		
*Sau*3AI	↓GATC		
*Sca*I	AGT↓ACT		

and the rare-cutting restriction endonuclease from *Saccharomyces cerevisiae* I-*Sce*I 18 bp:

5'-TAGGGATAACAGGGTAAT-3' (see SEQ ID NO: 1)

Plasmid constructions

Standard recombinant DNA techniques were used for cloning (Sambrook *et al.*; Molecular cloning - A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989)). In all cloning steps involving synthetic DNA linkers or PCR fragments, the correct DNA sequence of the linkers or PCR fragments was verified by DNA sequence analysis, using a Pharmacia LKB, ALF fluorescent sequencer.

Construction of the target site:

The **plasmid pUR5710** (see Figure 1) was constructed by cloning a 2.0 kb *Bam*HI/*Sal*I fragment containing a 5' part of the *pyrG* gene, which is present on the plasmid pAW4.1 (Gouka *et al.*; Curr. Genet. 27 (1995) 536-540), into the general cloning vector pIC20R (Marsh *et al.*; Gene 32 (1984) 481-485) digested with *Bam*HI and *Sal*I. Subsequently, a synthetic DNA linker containing the 18 bp recognition site for the I-*Sce*I endonuclease (5'-TAGGGATAACAGGGTAAT-3'; see SEQ ID NO: 1) flanked by *Sal*I and *Hin*dIII sites was cloned into the plasmid pUR5710 digested with *Sal*I and *Hin*dIII. This resulted in the **plasmid pUR5711** (see Figure 1). The **plasmid pUR5712** (see Figure 1) was constructed by cloning a 2.0 kb *Hin*dIII fragment containing sequences downstream of the *pyrG* coding region, which is present on the plasmid pAW4.4 (Gouka *et al.*; Curr. Genet. 27 (1995) 536-540), into the plasmid pUR5711 digested with *Hin*dIII. The orientation of this *Hin*dIII fragment compared to the coding region of the *pyrG* gene is identical to the wild-type situation. The plasmid pUR5712 was used to construct the *A. awamori* mutant *pyrG* strain AWCSCE.

Construction of the repair construct:

For the construction of the **plasmid pUR5713** (see Figure 2A) plasmid pAW4.4 was digested with *Hin*dIII, the *Hin*dIII site was filled in with Klenow and the fragment was subsequently digested with *Bam*HI. The resulting 1.6 kb fragment, containing sequences down-stream of the *pyrG* coding region,

was isolated. Furthermore, the plasmid pAW4.20 (Gouka *et al.*; Curr. Genet. 27 (1995) 536-540) was digested with *Bam*HI and *Hind*III and the 0.4 kb fragment, containing sequences present immediately upstream of the 1.6 kb fragment described above, was isolated. The 0.4 kb *Hind*III/*Bam*HI and 1.6 kb *Bam*HI/filled in *Hind*III fragments were simultaneously cloned into the general cloning vector pBluescript^R SK (Stratagene) digested with *Hind*III and *Sma*I. This resulted in the plasmid pUR5713. The **plasmid pUR5714** (see Figure 2B) was constructed by cloning a 1.0 kb *Bgl*II/*Hind*III fragment containing a 3' part of the *pyrG* gene, which is present on the vector pAW4.1, into the general cloning vector pBluescript^R SK digested with *Bam*HI and *Hind*III. The **cosmid pUR5716** (see Figure 3) is derived from the cosmid vector pJB8 (Ish-Horowicz,D. and Burke,J.F.; Nucleic Acids Res. 9 (1981) 2989) by replacing the *Eco*RI/*Hind*III polylinker fragment by a synthetic linker containing an *Eco*RI and a *Not*I restriction site having the following sequence:

(5'-AATTC AT GCGGCCGC T-3'
3'-G TA CGCCGGCG ATCGA-5' see SEQ ID NO: 2).

In this cloning step, the *Hind*III site is lost. The **cosmid pUR5718** (see Figure 3) was constructed by simultaneously cloning the 1.0 kb *Not*I/*Hind*III fragment from the plasmid pUR5714 and the 2.0 kb *Hind*III/*Not*I fragment from the plasmid pUR5713 into the plasmid pUR5716 digested with *Not*I. Thereby, this vector carries a sequence homologous to both sides of the I-*Sce*I site at the *pyrG* target locus in the *A. awamori* mutant *pyrG* strain AWCSCE.

Construction of multi-copy vectors:

The **plasmid pUR5729** (see Figure 4) was constructed by cloning the 1.5 kb *Pst*I/*Sac*I fragment containing the open reading frame (ORF) of the cutinase gene from *Fusarium solani pisi* (synthetic copy of the cDNA; Van Gemeren *et al.*; Journal of Biotechnology 40 (1995) 155-162) under control of the promoter and terminator of the *exlA* gene from *Aspergillus awamori* (Gouka *et al.*; Applied

Microbiology and Biotechnology 46, (1996) 28-35), from the plasmid pUR7385 (Van Gemeren *et al.*; Applied Microbiology and Biotechnology 45, (1996) 755-763), into the general cloning vector pIC19H (Marsh *et al.*; Gene 32 (1984) 481-485) digested with *Pst*I and *Sac*I. Based on the cosmid pUR5718 two new cosmids were constructed containing multiple copies of the cutinase gene under control of the *exlA* expression signals (as described above). A single copy of this expression cassette was isolated as a 1.5 kb *Hin*dIII fragment from the plasmid pUR5729 and ligated into the cosmid pUR5718 digested with *Hin*dIII. After transforming the ligation mixture into the *E. coli* strain DH5α, the **cosmid pUR5722** (see Figure 5) was obtained which contained a tandem array of four copies of the expression cassette. After packaging of the ligation mix using the λ-DNA in vitro packaging module (Amersham; code RPN1717), the packaging mix was transformed into *E. coli* strain 1046 (both according to the protocol provided with the module). From this transformation the **cosmid pUR5725** (see Figure 5) was obtained which contained a tandem array of nine copies of the expression cassette.

Construction of the I-*Sce*I expression vector:

The **plasmid pUR5736** (see Figure 6) was constructed by replacing a *Sca*I/*Bam*HI fragment containing a part of the promoter from the *A. nidulans gpd* gene fused to the coding region of the *E. coli hph* gene, which is present on the plasmid pAN7.1 (Punt *et al.*; Gene 56 (1987) 117-124), by a PCR fragment containing the same promoter fragment fused to a *Nde*I and *Bam*HI restriction site, digested with *Sca*I and *Bam*HI. To obtain the vector fragment from the plasmid pAN7.1 the plasmid was only partially digested with *Sca*I, because the pUC backbone of the plasmid contains another *Sca*I site. The PCR fragment was obtained in a PCR reaction on the plasmid pAN7.1 using the primers MGgpd1 (5'-GACAAGGTCG-TTGCGTCAGTC-3'; see SEQ ID NO: 3) and MGgpd2 (5'-CGGGATCCTT-CCATATGTGATGTCTGCTCAAGCGG-3'; see SEQ ID NO: 4). Subsequently, a *Bgl*II/*Hin*dIII fragment from the plasmid

pUR5736 containing the promoter from the *A. nidulans gpd* gene and the terminator sequences from the *A. nidulans trpC* gene, was cloned into the *Bam*HI/*Hin*dIII sites of the general cloning vector pBluescript^R SK. This resulted in the **plasmid pUR5737**. Hereafter, pUR5737 was digested with *Bam*HI, this site was filled in using the Klenow enzyme, and a second digestion with *Nde*I was performed. The plasmid pSCM525 (kindly provided by Prof. Dr. B. Dujon, Institut Pasteur, Paris and described in United State Patent no; 5,474,896; filed Nov.5, 1992), containing a synthetic gene encoding the I-*Sce*I endonuclease, was digested with *Sal*I, this site was filled in using the Klenow enzyme, and a second digestion with *Nde*I was performed. The resulting fragment was cloned into the plasmid pUR5737 (*Nde*I/filled in *Bam*HI fragment as described above) which resulted in the I-*Sce*I expression **vector pUR5724** (see Figure 7).

Transformation experiments

Preparation of protoplasts:

Conidia were obtained by growing the *A. awamori* strains at 30°C on a nitrocellulose filter (Hybond-N, Amersham) placed on a PDA (Potato Dextrose Agar) plate for several days and subsequently washing the filters with physiological salt solution.

Protoplasts of *A. awamori* were prepared as described by Punt and Van den Hondel (Methods in Enzymology 216 (1993) 447-457). A shake flask containing 200 ml of MM medium (0.4 ml 1 M $MgSO_4$, 2 ml 100 x spore elements (per litre; 60 g EDTA.$2H_2O$, 11 g $CaCl_2.2H_2O$, 7.5 g $FeSO_4.7H_2O$, 2.8 g $MnSO_4.H_2O$, 2.7 g $ZnSO_4.7H_2O$, 0.8 g $CuSO_4.5H_2O$, 0.9 g $CoCl_2.6H_2O$, 0.5 g $Na_2MoO_4.2H_2O$, 0.8 g H_3BO_3, 0.5 g KI, pH 4.0 with NaOH), 10 ml 20% glucose, 4 ml 50 x AspA (3.5 M $NaNO_3$, 0.35 M KCl, 0.55 M KH_2PO_4, pH 6.5 with KOH)) including 0.5% yeast extract was inoculated with 10^6 conidia/ml of *A. awamori* and incubated for 18 hours at 30°C in a shaker at 200 rpm. Mycelium was harvested through sterile Mirocloth^R and washed with ice-cold 0.6 M $MgSO_4$. The mycelium was resuspended in OM medium (per litre: 500 ml 2.4 M $MgSO_4$, 480 ml H_2O, 16.8 ml 0.5 M

Na_2HPO_4, 3.2 ml 0.5 M NaH_2PO_4, pH 5.8-5.9) at 5 ml/g mycelium. Subsequently, 5 mg Novozym 234^R and 6 mg BSA were added per g mycelium. Protoplasting was allowed to proceed for 1-2 hours at 30°C in a shaker at 80-100 rpm. The formation of protoplasts was checked using a light microscope. Protoplasts were filtered through sterile Miracloth^R and the sample was divided in 30 ml aliquots in falcon tubes. STC (1.2 M sorbitol, 10 mM Tris/HCl pH 7.5, 50 mM $CaCl_2.2H_2O$) was added to bring the volume up to 50 ml and the protoplasts were harvested by centrifugation at 2000 rpm for 10 minutes at 4°C. The protoplasts were washed again in 50 ml STC and resuspended in STC at a concentration of approximately 10^8 protoplasts/ml.

PEG transformations: Five to 7.5 µg of a single plasmid or two plasmids (in case the I-*Sce*I expression plasmid is co-transformed) was added to an aliquot of 100 µl (10^7) protoplasts, mixed and incubated for 25 minutes on ice. PEG was added in two 200 µl aliquots and an 850 µl aliquot, and the mixture was incubated at room temperature for 20 minutes. Finally, the mixture was washed with 10 ml of STC, harvested by centrifugation at 2000 rpm for 10 minutes at room temperature and the sample was plated on a MM plate for selection of transformants.

Construction of the *A. awamori* mutant *pyrG* strain AWCSCE: Transformation of the wild-type *A. awamori* strain was performed with a purified (Qiaex gel extraction kit; Qiagen cat. no. 20021) *Eco*RI fragment obtained from the plasmid pUR5712 containing the mutant *pyr*G gene with the I-*Sce*I restriction site at the site of the deletion (see figure 1 and 8). Per transformation 2 x 10^6 protoplasts were transformed with 10 µg of DNA. Since *pyrG* strains are resistant to 5-FOA (5-fluoro-orotic acid; Boeke *et al.* Mol. Gen. Genet. 197 (1984) 345-346), *pyrG* transformants can be selected directly from wild-type strains. Transformants were selected on MM plates (AspA is replaced by AspA-N; 50 x Aspa-N = 0.35 M KCl, 0.55 M KH_2PO_4, pH 6.5 with KOH)

supplemented with 10 mM Uridine and 0.75 mg/ml of 5-FOA, with 10 mM proline as the N-source. The mutant phenotype of the transformants that were obtained was checked by growing these colonies on MM plates without uridine. Two transformants that were not able to grow without uridine were further analyzed by Southern blot analysis (Figure 9 and see below).

DNA isolation, PCR and Southern analysis

Southern analysis was performed to confirm at a molecular level that the mould cell had been transformed and the desired DNA had been integrated into the genome.

To obtain mycelium material for a genomic DNA isolation, approximately 10^8 mould conidia were inoculated in 50 ml of *Aspergillus* minimal medium supplemented with 0.5% yeast extract and incubated for a period ranging from 22 hours to 3 days at 30 °C in a shaker at 200 rpm. The mycelium was harvested through Miracloth^R (Calbiochem) and snap frozen in liquid N_2. Frozen samples were ground to a fine powder using a Mikro-Dismembrator^R (ex Braun Biotech International) for 1 minute at 1750 rpm. Mould genomic DNA was isolated using Qiagen genomic tips (cat. no. 10223) and a protocol for genomic DNA purification from filamentous fungi provided by the supplier. The step for digestion of cell wall material was omitted.

The PCR reactions were performed in a Perkin Elmer DNA Thermal Cycler 480 using approximately 1 µg genomic DNA, 25 pMol of each primer, 10nMol of each dNTP, 1 unit of Taq DNA polymerase (Gibco-BRL) and 10 µl of 10 x Taq DNA polymerase buffer in a total volume of 100 µl. The reactions were overlaid with mineral oil. The amplification was started with 5 min at 94°C, followed by 30 cycles of 1 min 94°C, 1 min 55°C and 1 min 72°C. After the final cycle the elongation step was followed by another 5 min at 72°C. The sequence of the primers that were used are:

MGPyr1: 5'-GCCAGTACACTACTTCTTCG-3' (see SEQ ID NO: 5)

MGPyr2: 5'-AGGAGATCGCGAGAAGGTTG-3' (see SEQ ID NO: 6)

For the Southern blot, approximately 2.5 μg of DNA was digested with (a) restriction endonuclease(s) at 4 Units/μg for 16 hours. The following restriction endonucleases were used; *Sau*3AI, *Bgl*II, I-*Sce*I and *Sal*I. The DNA was separated on a 0.8% agarose TBE gel and transferred to a Hybond N membrane by capillary blotting (overnight). The membrane was (pre-)hybridized according to the Hybond protocol.

For the Southern blot presented in Figure 9 the chromosomal DNA (7.5 μg) was digested with *Sau*3AI. The blot was probed with an 18 bp end-labelled oligonucleotide representing the I-*Sce*I restriction site. This oligonucleotide was end-labelled using T4-polynucleotide kinase and γ-^{32}P-ATP. Hybridization was carried out for 4 hours at 42 °C. The filter was washed for 5 minutes with 2 x SSC at 42 °C followed by another 5 minutes with 2 x SSC, 0.1% SDS at 42 °C.

For the Southern blot presented in Figure 10 the chromosomal DNA was digested with *Bgl*II or *Bgl*II and I-*Sce*I. The 2.4 kb *Bam*HI x *Hin*dIII *pyrG* fragment from pAW4.1 was used as a probe. A DNA probe labelled with α^{32}P-dCTP was obtained using the RTS RadPrime DNA Labelling System from GibcoBRL (cat. no. 10387-017). The electronic autoradiographs were obtained using an Instant Imager (Packard).

Example 1A. Experimental setup.

The experimental design of the process for site-directed integration of multiple copies of a gene in the mould *A. awamori* using the I-*Sce*I endonuclease is shown in figure 8. The system is based on three components, a fungal strain containing the target sequence with a I-*Sce*I restriction site, a repair construct that is carrying sequences homologous to the target locus together with multiple copies of (a) gene(s) encoding (a) desired protein(s), and a plasmid containing an expression cassette with the gene encoding the I-*Sce*I endonuclease. In order to specifically detect integration by homologous recombination we used the

endogenous *pyrG* gene as a selectable marker gene (Gouka *et al*. Curr. Genet. 27 (1995) 536-540). First, we constructed a plasmid containing a defective *pyrG* gene in which a 0.8 kb region, encompassing the 3′ end of the coding region and flanking terminator sequences, was replaced by the I-*Sce*I restriction site (pUR5712). Using this plasmid the *A. awamori* mutant *pyrG* strain AWCSCE was constructed by a selection strategy for gene-replacement in fungi (see Materials and Methods). Second, the repair construct contains a complementary defective *pyrG* gene, in which 0.14 kb of the 5′ end of the coding region is deleted, that has sequences homologous to both sides of the I-*Sce*I site at the *pyrG* target locus in the *A. awamori* mutant *pyrG* strain AWCSCE. This repair construct contains an unique *Hin*dIII restriction site that can be used for inserting multiple copies of (a) gene(s) encoding (a) desired protein(s). The complete insert of the repair construct is flanked by *Not*I sites, which makes it possible to remove the vector sequences and transform *A. awamori* with only the insert fragment. Third, the expression cassette of the I-*Sce*I gene consists of the promoter from the efficiently expressed glyceraldehyde-3-phosphate dehydroge-nase encoding gene from *Aspergillus nidulans* (*gpdA*) (Punt *et al*. Gene 56 (1987) 117-124), an artificial I-*Sce*I ORF (United States Patent no; 5,474,896; filed Nov.5, 1992) and the transcription termination region of the *A. nidulans trpC* gene (Mullaney *et al*. Mol. Gen. Genet. 199 (1985) 37-45). When the expression plasmid and the repair construct are co-introduced into protoplasts of the strain AWCSCE, transient expression of the I-*Sce*I gene may lead to the introduction of double-strand breaks at the I-*Sce*I site, thereby stimulating homologous recombination with the repair construct and integration of the multiple gene copies. Alternatively, the I-*Sce*I endonuclease may be introduced directly into the cell in a way that is similar to Restriction Mediated Integration (REMI; Kuspa and Loomis, Proc. Natl. Acad. Sci. USA 89 (1992) 8803-8807; Redman and Rodriguez, Exp. Mycol. 18 (1994) 230-246;

Brenneman *et al*. Proc. Natl. Acad. Sci. USA 93 (1996) 3608-3612). Because homologous recombination will restore an intact *pyrG* gene, these events can be selected directly by growing the transformed cells on MM plates without uridine.

Example 1B. Determination of the occurrence of a natural I-*Sce*I restriction site in the genome of *A. awamori*.

Before the system of this invention was set-up, we determined whether naturally occurring I-*Sce*I restriction sites are present in the genome of *A. awamori*. The I-*Sce*I endonuclease has an 18 bp recognition site, which will statistically occur only once in 6.9×10^{10} bp (4^{18}) or approximately once in every 18.500 *A. awamori* genomes. Therefore it seems unlikely that an I-*Sce*I site will be present in the genome.

In order to determine the presence of a naturally occurring I-*Sce*I site, a Southern blot was performed with *A. awamori* genomic DNA. The genomic DNA was digested with *Sau*3AI, which does not cut within the I-*Sce*I restriction site, to create smaller fragments that are more suitable for Southern blotting and hybridization with a labelled oligonucleotide. Plasmid reconstructions with the plasmids pUR5712 and pSCM522 (control DNA substrate containing the I-*Sce*I restriction site, supplied with the I-*Sce*I endonuclease from Boehringer Mannheim, cat. no. 1497235), both containing a single I-*Sce*I restriction site, representing a single copy site or 20 or 200 sites per genome were included as controls. The blot was probed with an 18 bp end-labelled oligonucleotide representing the I-*Sce*I restriction site. For the autoradiograph see Figure 9. Whereas the single copy reconstructions with the control plasmids show a clear hybridizing fragment (lanes 3 and 6), no hybridizing fragments are present in the lanes 4 and 5 containing the chromosomal *A. awamori* DNA. This result demonstrates that the genome of *A. awamori* does not contain a natural I-*Sce*I restriction site. Thus, it is possible to specifically engineer an unique I-*Sce*I site into the genome

at a locus of choice and subsequently introduce an unique double-strand break in the genomic DNA at that locus by expressing the I-*Sce*I gene in the cell or introducing the I-*Sce*I endonuclease itself.

Example 1C. Construction of the *A. awamori* mutant *pyrG* strain AWCSCE which contains an I-*Sce*I restriction site at the locus of the mutated *pyrG* gene.

The *A. awamori* mutant *pyrG* strain AWCSCE is obtained by replacing the chromosomal wild-type *pyrG* gene with the mutant *pyrG* gene from pUR5712, which contains the I-*Sce*I restriction site at the site of the deletion. Therefore, the *A. awamori* strain was transformed with a purified *Eco*RI fragment containing the insert from the plasmid pUR5712. Per transformation 2 x 10^6 protoplasts were transformed with 10 μg of DNA. In seven transformations a total of 11 5-FOA^R colonies were obtained. All these transformants were not able to grow on MM plates without uridine, corresponding with a *pyrG* phenotype. Genomic DNA was isolated from these strains and a PCR was performed with the primers MGPyr1 and MGPyr2. These primers anneal within the regions flanking the deletion and will generate a 1.13 kb fragment when the wild-type *pyrG* gene is present and a 0.36 kb fragment when the mutant *pyrG* gene is present. Two out of the 11 5-FOA^R colonies contained the 0.36 kb fragment specific for the mutant *pyrG* gene and were further analyzed by Southern blotting. The remaining 9 5-FOA^R colonies are likely to be the result of spontaneous mutations. DNA was digested with *Bgl*II and I-*Sce*I and the blot was probed with a 2.4 kb *Bam*HI x *Hin*dIII fragment from pAW4.1 containing the *A. awamori pyrG* gene (Figure 10). In the DNA of both mutant strains and the wild-type strain a 9.0 kb *Bgl*II fragment was present. The other 2.7 kb *Bgl*II fragment, as present in the wild-type strain (lane 3), is absent in the mutant strains and replaced by a 0.5 kb *Bgl*II x I-*Sce*I fragment characteristic for the mutant *pyrG* gene carrying the I-*Sce*I restriction site at the site of the deletion. This result indicates that these mutants

originate from a replacement of the wild-type *pyrG* gene by the mutated *pyrG* gene from pUR5712. Because the mutant strain from lane 1 contains an additional unexplained hybridizing fragment of approximately 2.0 kb, the mutant strain from lane 2 was chosen for further experiments and designated AWCSCE.

Example 1D. Induction of site-directed integration at the *pyrG* locus by I-*Sce*I expression.

Protoplasts derived from the *A. awamori* mutant *pyrG* strain AWCSCE were transformed with a repair construct in the presence or absence of the I-*Sce*I expression vector pUR5724. The repair constructs consisted of a mutant complementing *pyrG* gene (pUR5718), and derivatives thereof containing a repeat of 4 copies (pUR5722) or 9 copies (pUR5725) of a cutinase expression cassette. Prior to transformation, DNA from the repair constructs was digested with *Not*I. This released the insert from the vector sequences, thereby creating ends that are homologous to the target locus. The results of the transformation experiments are shown in table 1. The transformation frequencies were calculated from parallel transformations with the positive control construct pAW4.2, containing the wild-type *pyrG* gene.

Transformation of pUR5718 without the I-*Sce*I expression vector pUR5724 yielded 16 recombinants corresponding to a gene targeting frequency of 10.6%, whereas including the I-*Sce*I expression vector yielded 63 recombinants corresponding to a gene targeting frequency of 41.7%. These results indicate that gene targeting with the repair construct pUR5718 is stimulated approximately four-fold by the introduction of a double strand break at the *pyrG* locus in the strain AWCSCE.

Transformation of pUR5722, containing the repeat of 4 copies of the cutinase expression cassette, without the I-*Sce*I expression vector pUR5724 yielded no recombinants. This means that the gene targeting frequency is lower than 3.6%. In contrast, transformation of pUR5722 with the

I-*Sce*I expression vector pUR5724 yielded 5 recombinants correspon-ding to a gene targeting frequency of 18%. These results indicate that gene targeting of a repair construct containing a tandem array of 4 copies of the cutinase gene is about 2 to 3 fold less efficient than that of the repair construct pUR5718 which only contains *pyrG* sequences. Moreover, these results indicate that gene targeting with the repair construct pUR5722 is stimulated at least five-fold by the introduction of a double strand break at the *pyrG* locus in the strain **AWCSCE**.

Transformation of pUR5725, containing the repeat of 9 copies of the cutinase expression cassette, without the I-*Sce*I expression vector pUR5724 yielded also no recombinants. This means that in this case the gene targeting frequency is lower than 0.6%. In contrast, transformation of pUR5725 with the I-*Sce*I expression vector pUR5724 yielded 2 recombinants corresponding to a gene targeting frequency of only 0.4%. These results indicate that gene targeting of a repair construct containing a tandem array of 9 copies of the cutinase gene is about 20 to 100 fold less efficient than that of the repair construct pUR5718 which only contains *pyrG* sequences.

In conclusion, these results demonstrated that gene targeting with the repair construct containing multiple gene copies of a heterologous gene was only possible when a double strand break was introduced at the *pyrG* locus in the strain AWCSCE. Gene targeting of a repair DNA becomes much more inefficient when an increasing number of gene copies are included with the repair construct. This further confirms the problem of site-directed integration of multiple gene copies as discussed in the section background of the invention and prior art.

Example 1E. Southern blot analysis of recombination events

The wild-type *pyrG* phenotype of several recombinants was confirmed by streaking the conidia on MM plates. This was done for 9 and 10 recombinants obtained from

transformations with pUR5718 in the absence or presence of the I-*Sce*I, respectively, and all the recombinants obtained from transformations with pUR5722 and pUR5725. From ten of these transformants, conidia from individual colonies were streaked again on MM plates. Subsequently conidia were isolated and cultures were grown to obtain mycelium for genomic DNA isolation. DNA isolation and Southern analysis is described in Materials and Methods. The genomic DNA was digested with *Bgl*II or *Sal*I. The Southern blots were probed with either the 2.4 kb *Bam*HI x *Hin*dIII *pyrG* fragment from pAW4.1, the 0.46 kb *Afl*II x *Sac*I fragment encompassing the terminator region from the endoxylanase gene (pUR5729), the 0.72 kb *Bam*HI x *Sal*I fragment encompassing the I-*Sce*I gene or the pJB8 vector. The experimental setup of the Southern analysis is shown in Figure 11 and the autoradiographs of the Southern blots hybridized with the *pyrG* and *TexlA* probes are depicted in Figure 12A and B, respectively. In the Southern blot of wild-type *A. awamori* genomic DNA digested with *Sal*I and probed with the 2.4 kb *pyrG* fragment a 3.3 kb and 3.8 kb fragment are present (Figure 12A, lane 4). In the mutant strain AWCSCE the 3.8 kb fragment is replaced by a 3 kb fragment (Figure 12A, lane 5). In recombinants obtained with the plasmid pUR5718 restoration of an intact *pyrG* gene will lead to a replacement of the 3 kb fragment with a 3.7 kb fragment (Figure 12A, lane 1-3). The small difference in size of the latter fragment compared to the same fragment in the wild-type strain is caused by a small 0.15 kb marker deletion in the 3' flanking sequence of the *pyrG* gene in the repair construct pUR5718. Because there is no *Sal*I site present within the cutinase expression cassette, site-directed integration of multiple cutinase gene copies in recombinants obtained with the plasmids pUR5722 (four cutinase gene copies) and pUR5725 (nine cutinase gene copies) is expected to lead to a replacement of the 3 kb fragment with a 9.6 kb or a 17.1 kb fragment, respectively. This replacement is observed for one pUR5725 recombinant (lane 7) and one pUR5722 recombinant (lane 10), which demonstrates the successful

one step site-directed integration of multiple gene copies in *A. awamori*. Unexpectedly, in the other recombinants one or more copies of the cutinase gene have been lost during recombination. The remaining pUR5722 recombinants contain two copies of the cutinase gene (Figure 12A, lane 9, 11 and 12) or only one copy (Figure 12A, lane 8). The other pUR5725 recombinant (Figure 12A, lane 6) contains three copies of the cutinase gene.

Figure 12B depicts the Southern blot of genomic DNA digested with *Bgl*II and probed with the *TexlA* probe. This probe hybridizes with the endogenous *exlA* gene and the introduced multiple copies of the cutinase expression cassette. Thus the wild-type strain, the mutant AWCSCE strain and the recombinants obtained with pUR5718 contain only a 6 kb *Bgl*II fragment representing the endogenous *exlA* gene (lanes 1-5). Because the cutinase expression cassette contains one *Bgl*II site, this digestion will generate a 1.5 kb fragment in the recombinants containing multiple copies of the cutinase gene. The intensity of this repeat fragment relative to the endogenous *exlA* fragment is an indication for the number of cutinase gene copies that are present. The recombinants obtained with pUR5725 (lane 6 and 7) contain the additional 1.5 kb fragment. The intensities of the bands correspond well to the presence of three (lane 6) or nine copies (lane 7) of the cutinase gene as was also determined by the total size of the fragments (Figure 12A). The same is true for the recombinants obtained with pUR5722. Due to the orientation of the cutinase gene copies relative to the *pyrG* gene (see figure 5) the *Bgl*II digestion will generate the 1.5 kb repeat fragment and a 2 kb border fragment. The recombinant in lane 8 contains only the repeat fragment, indicating the presence of one copy of the cutinase gene. The recombinants in lanes 9, 11 and 12 contain a repeat fragment that has the same intensity as the border fragment, which confirms the presence of two copies of the cutinase gene. The recombinant is lane 10 contains a repeat fragment that is about three times more

intense than the border fragment which confirms the presence of four copies of the cutinase gene.
In order to determine if other DNA, such as cosmid vector sequences, or the plasmid containing the I-*Sce*I expression cassette had co-integrated into the genome in the recombinant lines, the Southern blots have also been probed with the 0.72 kb *Bam*HI x *Sal*I fragment encompassing the I-*Sce*I gene (*Bgl*II digested DNA) or the pJB8 vector (*Sal*I digested DNA). These blots demonstrated that in none of the recombinants, except for the pUR5725 recombinant containing nine copies of the cutinase gene, other DNA had been integrated (results not shown). This demonstrates that it is possible to construct "food-grade" strains that contain multiple copies of a gene without the presence of other foreign DNA. It should be noted that in the experiments described here the inserts of pUR5722 and pUR5725 had not been purified from the vector DNA prior to transformation, whereas this is possible. Moreover, it may be possible to omit the use of the I-*Sce*I expression vector by transforming the endonuclease directly. These modifications will further improve the selection of strains that do not contain other foreign DNA.

This Example shows that advantages of this process for site-directed integration of multiple copies of a gene in a mould are:

- multiple gene copies can be introduced at a predetermined locus in the genome;
- the possibility to obtain site-directed integration of multiple gene copies in the genome is significantly improved by the introduction of a specific double-strand break at the chromosomal target in the mould cell; and
- the method results in a mould strain without residues of bacterial antibiotic resistance markers or other bacterial sequences like origins of replication, which the consequence that the resulting mould

strains or products derived therefrom are so-called "food-grade" products.

REFERENCES

- Archer *et al.* Antonie van Leeuwenhoek 65 (1994) 245-250
- Boeke *et al.* Mol. Gen. Genet. 197 (1984) 345-346
- Brenneman *et al.* Proc. Natl. Acad. Sci. USA 93 (1996) 3608-3612
- Cami, B. and Kourilsky,P.; Nucl. Acids Res. 5 (1978) 2381
- Choulika *et al.* Mol. Cell. Biol. 15 (1995) 1968-1973
- Choulika *et al.* International Application published under the Patent Cooperation Treaty, WO 96/14408 (1996)
- Colleaux *et al.* Cell 44 (1986) 521-533
- Colleaux *et al.* Proc. Natl. Acad. Sci. USA 85 (1988) 6022-6026
- Dujon *et al.*, Institute Pasteur, Paris and described in United State Patent no; 5,474,896; filed Nov.5, 1992
- De Massy and Nicolas, EMBO-J 12 (1993) 1459-1466)
- Gimble and Thorner, Nature 357 (1992) 301-306
- Gouka *et al.* Applied and Environmental Microbiology 62 (1996) 1951-1957
- Gouka *et al.*; Applied Microbiology and Biotechnology 46, (1996) 28-35
- Gouka *et al.*; Curr. Genet. 27 (1995) 536-540
- Halfter *et al.* Mol. Gen. Genet. 231 (1992) 186-193
- Hanahan; J. Mol. Biol. 166 (1983) 557-580
- Hamilton *et al.*; Proc. Natl. Acad. Sci. USA 93 (1996) 9975-9979
- Ish-Horowicz,D. and Burke, J.F.; Nucleic Acids Res 9 (1981) 2989
- Jasin, TIG 12 (1996) 224-228
- Kucherlapati, R. and G.R. Smith (1988) editors, Genetic recombination, Washington DC: American Society of Microbiology
- Kuspa and Loomis, Proc. Natl. Acad. Sci. USA 89 (1992) 8803-8807
- Lee *et al.* Plant Cell 2 (1990) 415-425
- Lin *et al.*, Mol. Cell. Biol. 4 (1984) 1020-1034

- Marsh *et al.*; Gene 32 (1984) 481-485
- Meselson and Radding, Proc. Natl. Acad. Sci. USA 72 (1975) 358-361
- Mullaney *et al.* Mol. Gen. Genet. 199 (1985) 37-45
- Nicolas *et al.* Nature 338 (1989) 35-39
- Offringa *et al.* EMBO-J 9 (1990) 3077-3084
- Paszkowski *et al.* EMBO-J 7 (1988) 4021-4026
- Puchta *et al.* Nucleic Acids Research 21 (1993) 5034-5040
- Puchta *et al.* Proc. Natl. Acad. Sci. USA 93 (1996) 5055-5060
- Punt *et al.*; Gene 56 (1987) 117-124
- Punt and Van den Hondel; Methods in Enzymology 216 (1993) 447-457
- Redman and Rodriguez, Exp. Mycol. 18 (1994) 230-246
- Rouet *et al.* Proc. Natl. Acad. Sci. USA 91 (1994) 6064-6068
- Rouet *et al.* Mol. Cell. Biol. 14 (1994) 8096-8106
- Royer *et al.*; Bio/Technology 13 (1995) 1479-1483
- Sambrook *et al.*; Molecular cloning - A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989)
- Sargent *et al.* Mol. Cell. Biol. 17 (1997) 267-277
- Smith *et al.* Nucleic Acids Research 24 (1995) 5012-5019
- Strathern *et al.* Cell 31 (1982) 183-192
- Sun *et al.* Nature 338 (1989) 87-90
- Sun *et al.* Cell 64 (1991) 1155-1161
- Szostak *et al.*, Cell 33 breaks (1983) 25-35
- Thomas and Cappechi, Nature 346 (1990) 847-850
- Timberlake, "Gene Cloning and Analysis" (Chapter 3) in the book "More Gene Manipulations in Fungi" (1991) 51-85, edited by Bennett and Lasure
- Van den Hondel *et al.*; "Heterologous gene expression in filamentous fungi" (Chapter 18) in the book "More Gene Manipulations in Fungi" (1991) 397-428, edited by Bennett and Lasure
- Van den Hondel and Punt, "Gene transfer systems and vector development" (Chapter 1) in the book "Applied Molecular Genetics" (1991) 1-28, edited by Peberdy *et al.*

- Van den Hondel *et al.* Antonie van Leeuwenhoek 61 (1992) 153-160
- Van Gemeren *et al.*; Journal of Biotechnology 40 (1995) 155-162
- Van Gemeren *et al.*; Applied Microbiology and Biotechnology 45, (1996) 755-763
- Van Gemeren, "Expression and secretion of defined cutinase variants by *Aspergillus awamori*" (Chapter 5) Thesis University of Utrecht (1997) ISBN 90-393-1229-X
- Verdoes *et al.* Transgenic Research 2 (1993) 84-92
- Verdoes, "Strain improvement strategies for the overproduction of fungal proteins by filamentous fungi" (Chapter 1) Thesis Free University Amsterdam (1994)
- Verdoes *et al.*; Appl. Microbiol. Biotechnol. 43 (1995) 195-205
- Wenzlau *et al.* Cell 56 (1989) 421-430
- Zimmer and Gruss, Nature 338 (1989) 150-153
- Zin and Butow, Cell 40 (1985) 887-895

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT (except for U.S.A.):

(A) NAME: Unilever N.V.

(B) STREET: Weena 455

(C) CITY: Rotterdam

(E) COUNTRY: The Netherlands

(F) POSTAL CODE (ZIP): NL-3013 AL

(i) APPLICANT (except for U.S.A.):

(A) NAME: Unilever PLC

(B) STREET: Unilever House, Blackfriars

(C) CITY: London

(E) COUNTRY: United Kingdom

(F) POSTAL CODE (ZIP): EC4P 4BQ (GB)

(i) APPLICANT (for U.S.A. only):

(A) NAME: Marcellus Johannes Augustinus DE GROOT
c/o Unilever Research Vlaardingen

(B) STREET: Olivier van Noortlaan 120

(C) CITY: Vlaardingen

(E) COUNTRY: The Netherlands

(F) POSTAL CODE (ZIP): NL-3133 AT

(i) APPLICANT (for U.S.A. only):

(A) NAME: Alida Godelieve Maria BEIJERSBERGEN
c/o Unilever Research Vlaardingen

(B) STREET: Olivier van Noortlaan 120

(C) CITY: Vlaardingen

(E) COUNTRY: The Netherlands

(F) POSTAL CODE (ZIP): NL-3133 AT

(i) APPLICANT (for U.S.A. only):

(A) NAME: Wouter MUSTERS
c/o Unilever Research Vlaardingen

(B) STREET: Olivier van Noortlaan 120

(C) CITY: Vlaardingen

(E) COUNTRY: The Netherlands

(F) POSTAL CODE (ZIP): NL-3133 AT

(ii) TITLE OF INVENTION: A process for site-directed integration of multiple copies of a gene in a mould

(iii) NUMBER OF SEQUENCES: 4

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER: ..... (not yet known)

(2) INFORMATION FOR SEQ ID NO: 1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"
(vii) IMMEDIATE SOURCE:
(B) CLONE: restriction site for I-*Sce*I endonuclease from *Saccharomyces cerevisiae*
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
TAGGGATAAC AGGGTAAT 18

(2) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(vii) IMMEDIATE SOURCE:

(B) CLONE: synthetic linker containing *Eco*RI and *Not*I restriction sites replacing an *Eco*RI/*Hind*III polylinker fragment thereby destroying the *Hind*III site

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

AATTCATGCG GCCGCTAGCT 20

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "synthetic DNA"

(vii) IMMEDIATE SOURCE:

(B) CLONE: primer MGgpd1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GACAAGGTCG TTGCGTCAGT C 21

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "synthetic DNA"

(vii) IMMEDIATE SOURCE:

(B) CLONE: primer MGgpd2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CGGGATCCTT CCATATGTGA TGTCTGCTCA AGCGG 35

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "synthetic DNA"

(vii) IMMEDIATE SOURCE:

(B) CLONE: primer MGPyr1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GCCAGTACAC TACTTCTTCG 20

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "synthetic DNA"

(vii) IMMEDIATE SOURCE:

(B) CLONE: primer MGPyr2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AGGAGATCGC GAGAAGGTTG 20

C L A I M S

1. A process for site-directed integration of multiple copies of a gene in a mould, which comprises

(i) providing a mould cell containing in its chromosomal DNA a restriction site for a rare-cutting endonuclease,

(ii) transforming such mould cell with a piece of DNA comprising in the 5′ to 3′ direction in the following order

(a) a first DNA fragment homologous to part of the DNA upstream and in the neighbourhood of the restriction site for the rare-cutting endonuclease present in the chromosomal DNA of the mould

(b) multiple copies of at least one expressible gene comprising a structural gene encoding a desired protein,

(c) a second DNA fragment homologous to part of the DNA downstream and in the neighbourhood of the restriction site for the rare-cutting endonuclease present in the chromosomal DNA of the mould,

while during the transformation of the mould the presence of the rare-cutting endonuclease is provided,

(iii) selecting or screening for a mould cell in which the multiple gene copies of said expressible gene are inserted into the chromosomal DNA of the mould.

2. A process according to claim 1, in which the rare-cutting endonuclease is I-*Sce*I.

3. A process according to claim 1, in which the restriction site for the rare-cutting endonuclease has been introduced at a desired locus.

4. A process according to claim 3, in which the desired locus is within a selectable marker gene or in the neighbourhood thereof.

5. A process according to claim 4, in which the piece of DNA comprises a third DNA fragment that completes any disrupted or partially deleted selectable marker gene in the chromosomal DNA.

6. A process according to claim 4, in which the part of the DNA **up-stream** of the restriction site for the rare-cutting endonuclease present in the chromosomal DNA of the mould, to which the **first** DNA fragment is homologous, is part of a selectable marker gene.

7. A process according to claim 4, in which the part of the DNA **down-stream** of the restriction site for the rare-cutting endonuclease present in the chromosomal DNA of the mould, to which the **second** DNA fragment is homologous, is part of a selectable marker gene.

8. A process according to claim 1, in which the restriction site for the rare-cutting endonuclease occurs naturally in the chromosal DNA of the mould.

9. A process according to claim 1, in which two or more restriction sites for the rare-cutting endonuclease are present.

10. A process according to claim 1, in which the expressible gene comprises (1) a promoter operable in said mould, (2) optionally a DNA fragment encoding a secretion signal peptide facilitating the secretion of said desired protein from said mould, (3) a structural gene encoding said desired protein, and (4) optionally a terminator operable in said mould, whereby the promoter and the optional terminator control the expression of the structural gene.

11. A process according to claim 1, in which during the transformation of the mould the rare-cutting endonuclease is provided by adding the endonuclease as such, and/or is formed *in situ* by co-transforming the mould with DNA encoding the rare-cutting endonuclease, which DNA is to be expressed during or after the transformation of the mould.

12. A process according to claim 1, in which the mould belongs to the group of *Eumycota*, and preferably is selected from the group consisting of the fungal sub-divisions *Ascomycotina*, *Basidiomycotina*, *Deuteromycotina*, *Mastigomycotina*, and *Zygomycotina*.

13. A process according to claim 12, in which the mould is selected from the genus *Aspergillus*, and preferably belongs to the species *Aspergillus awamori*.

14. A transformed mould obtainable by a process as claimed in claim 1.

15. A process for culturing a transformed mould obtained by a process as claimed in claim 1 or obtainable by such process.

16. A process for producing and optionally secreting a desired protein by carrying out a process as claimed in claim 15 under conditions whereby the structural gene encoding said desired protein is expressed, and optionally isolating or concentrating the desired protein.

Figure 1.

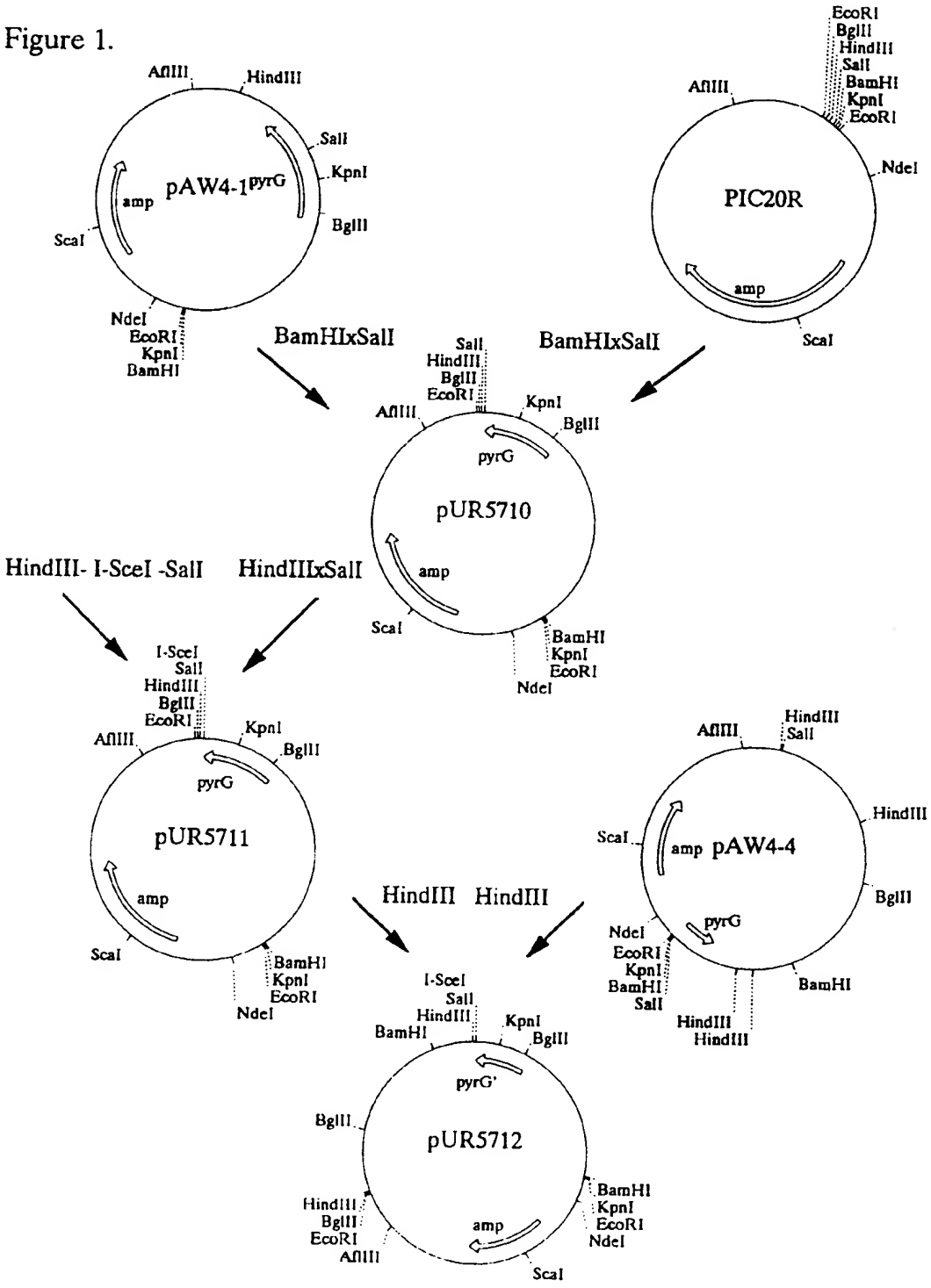
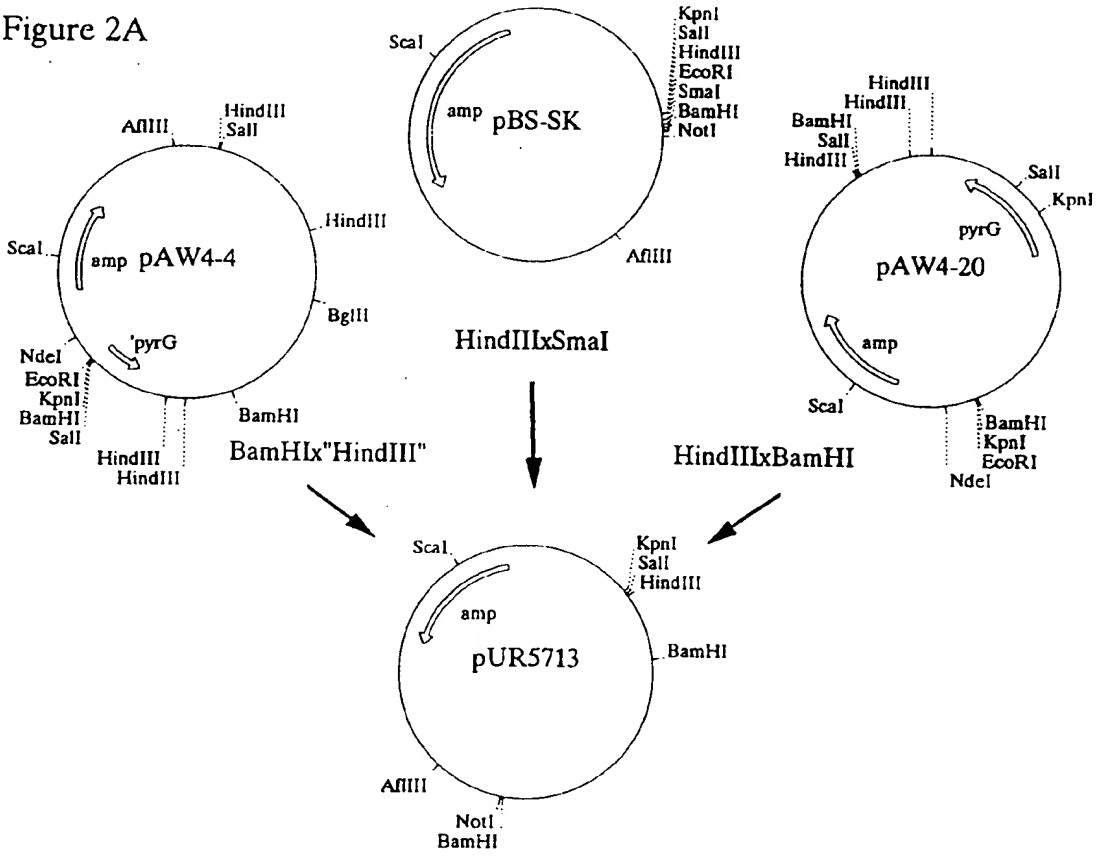

AflIII
HindIII
SalI
KpnI
BglII
ScaI
NdeI
EcoRI
BamHI
pAW4-1
pyrG
amp
PIC20R
BamHIxSalI
pUR5710
HindIII- I-SceI -SalI
HindIIIxSalI
pUR5711
HindIII HindIII
pAW4-4
pUR5712
pyrG'

Figure 2A


ScaI
amp
pBS-SK
KpnI
SalI
HindIII
EcoRI
SmaI
BamHI
NotI
AflIII
AflIII
HindIII
SalI
ScaI
amp
pAW4-4
HindIII
BglII
'pyrG
NdeI
EcoRI
KpnI
BamHI
SalI
BamHI
HindIII
HindIII
HindIIIxSmaI
BamHIx"HindIII"
HindIIIxBamHI
HindIII
HindIII
BamHI
SalI
HindIII
pyrG
SalI
KpnI
pAW4-20
amp
ScaI
BamHI
KpnI
EcoRI
NdeI
ScaI
amp
pUR5713
KpnI
SalI
HindIII
BamHI
AflIII
NotI
BamHI

2B

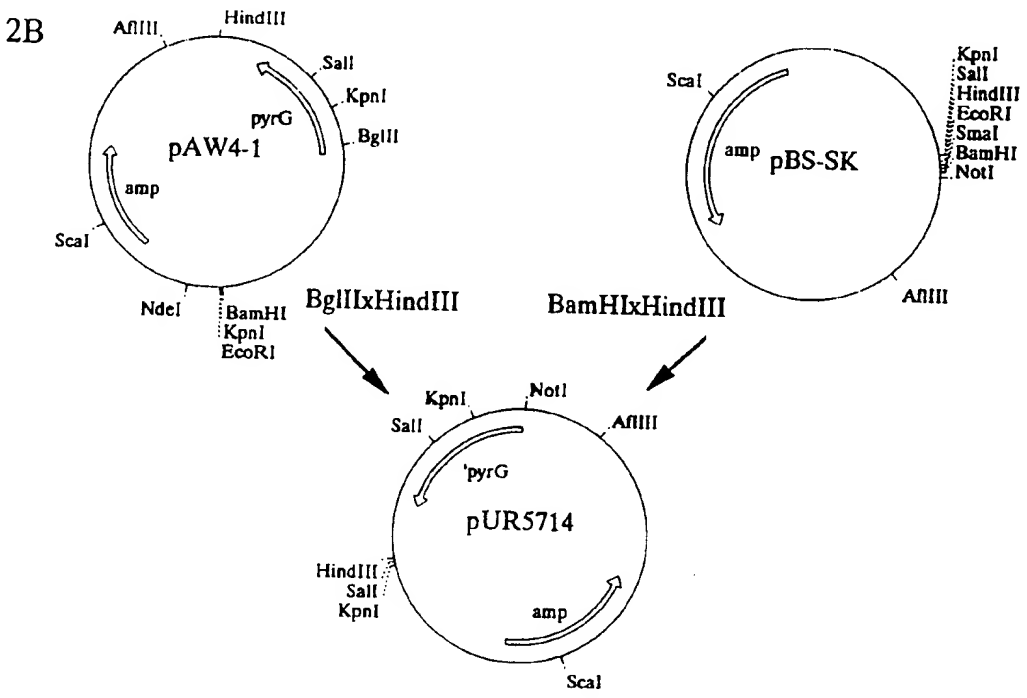

AflIII
HindIII
SalI
KpnI
BglII
pyrG
pAW4-1
amp
ScaI
NdeI
BamHI
KpnI
EcoRI
BglIIxHindIII
BamHIxHindIII
ScaI
KpnI
SalI
HindIII
EcoRI
SmaI
BamHI
NotI
amp
pBS-SK
AflIII
KpnI
NotI
SalI
AflIII
'pyrG
pUR5714
HindIII
SalI
KpnI
amp
ScaI

Figure 3.

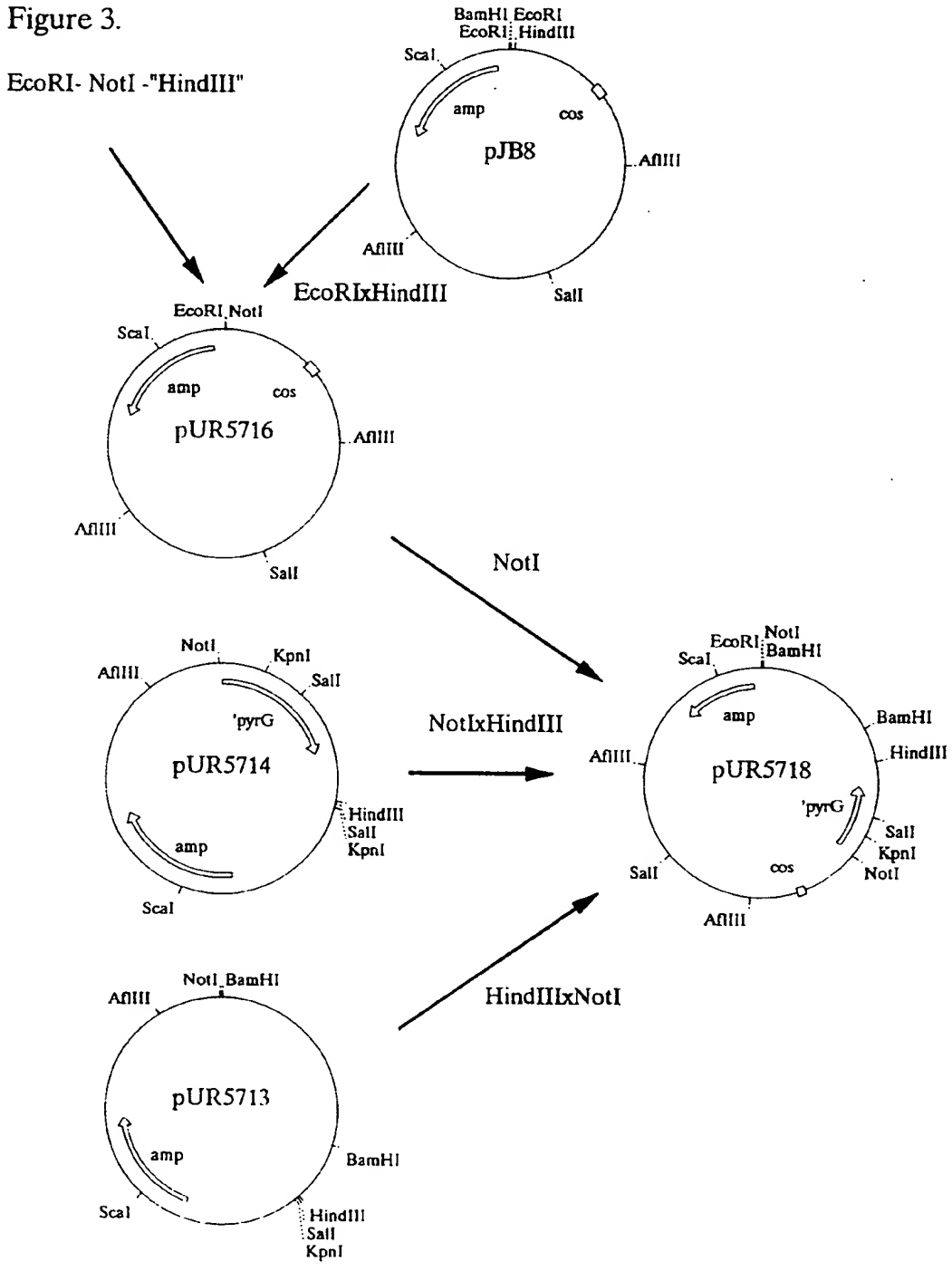

EcoRI- NotI -"HindIII"
BamHI EcoRI
EcoRI HindIII
ScaI
amp
cos
pJB8
AflIII
AflIII
SalI
EcoRIxHindIII
EcoRI NotI
ScaI
amp
cos
pUR5716
AflIII
AflIII
SalI
NotI
NotI
KpnI
AflIII
SalI
'pyrG
pUR5714
NotIxHindIII
HindIII
SalI
KpnI
amp
ScaI
EcoRI NotI
BamHI
ScaI
amp
BamHI
HindIII
AflIII
pUR5718
'pyrG
SalI
KpnI
SalI
cos
NotI
AflIII
NotI BamHI
AflIII
pUR5713
HindIIIxNotI
amp
BamHI
ScaI
HindIII
SalI
KpnI

Figure 4.

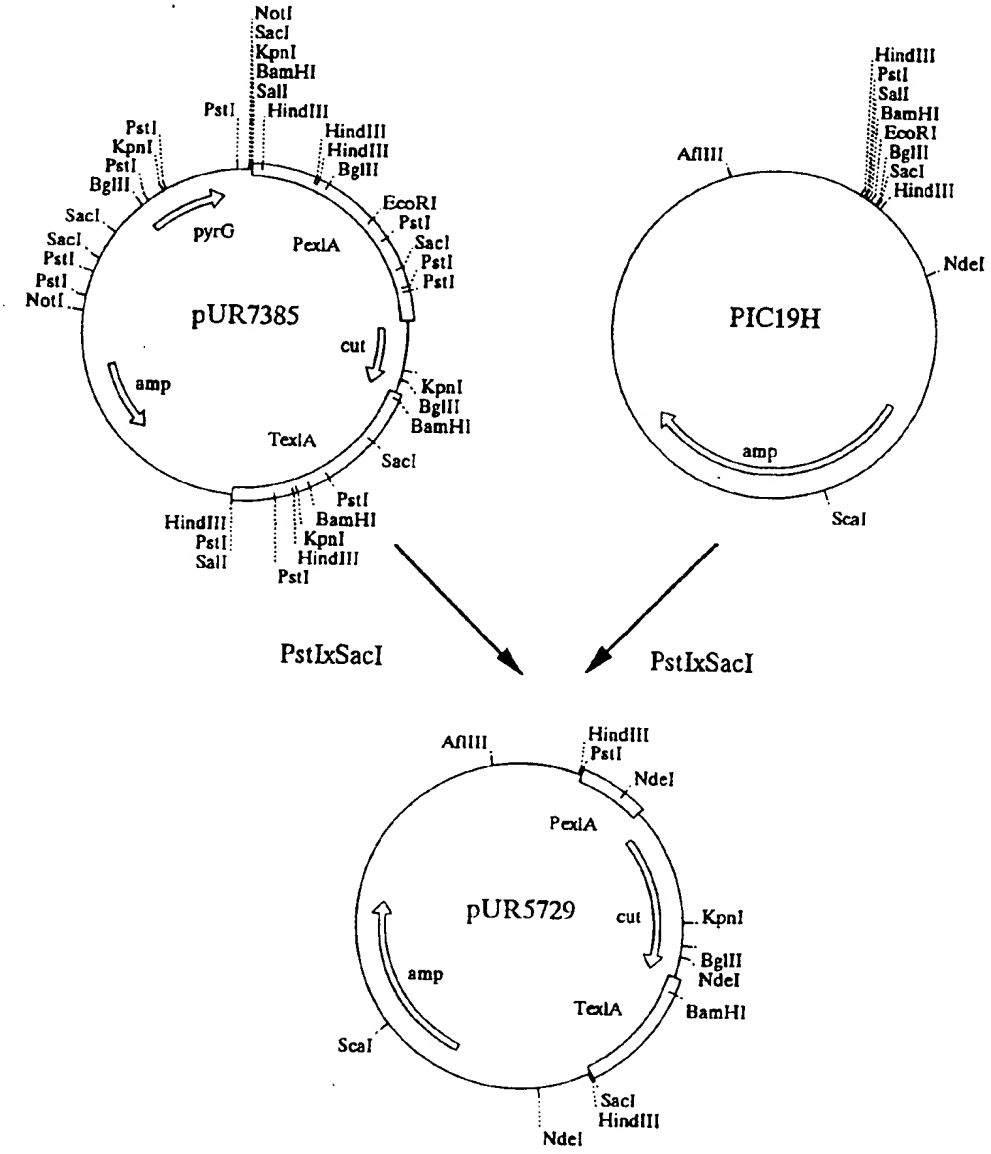

NotI
SacI
KpnI
BamHI
SalI
PstI
HindIII
HindIII
HindIII
BglII
EcoRI
PstI
SacI
PstI
PstI
pyrG
PexlA
pUR7385
cut
amp
KpnI
BglII
BamHI
TexlA
SacI
PstI
BamHI
KpnI
HindIII
PstI
HindIII
PstI
SalI
PstI
KpnI
PstI
BglII
SacI
SacI
PstI
PstI
NotI
AflIII
HindIII
PstI
SalI
BamHI
EcoRI
BglII
SacI
HindIII
NdeI
PIC19H
amp
ScaI
PstIxSacI
PstIxSacI
AflIII
HindIII
PstI
NdeI
PexlA
pUR5729
cut
KpnI
BglII
NdeI
BamHI
TexlA
amp
ScaI
SacI
HindIII
NdeI

Figure 5.

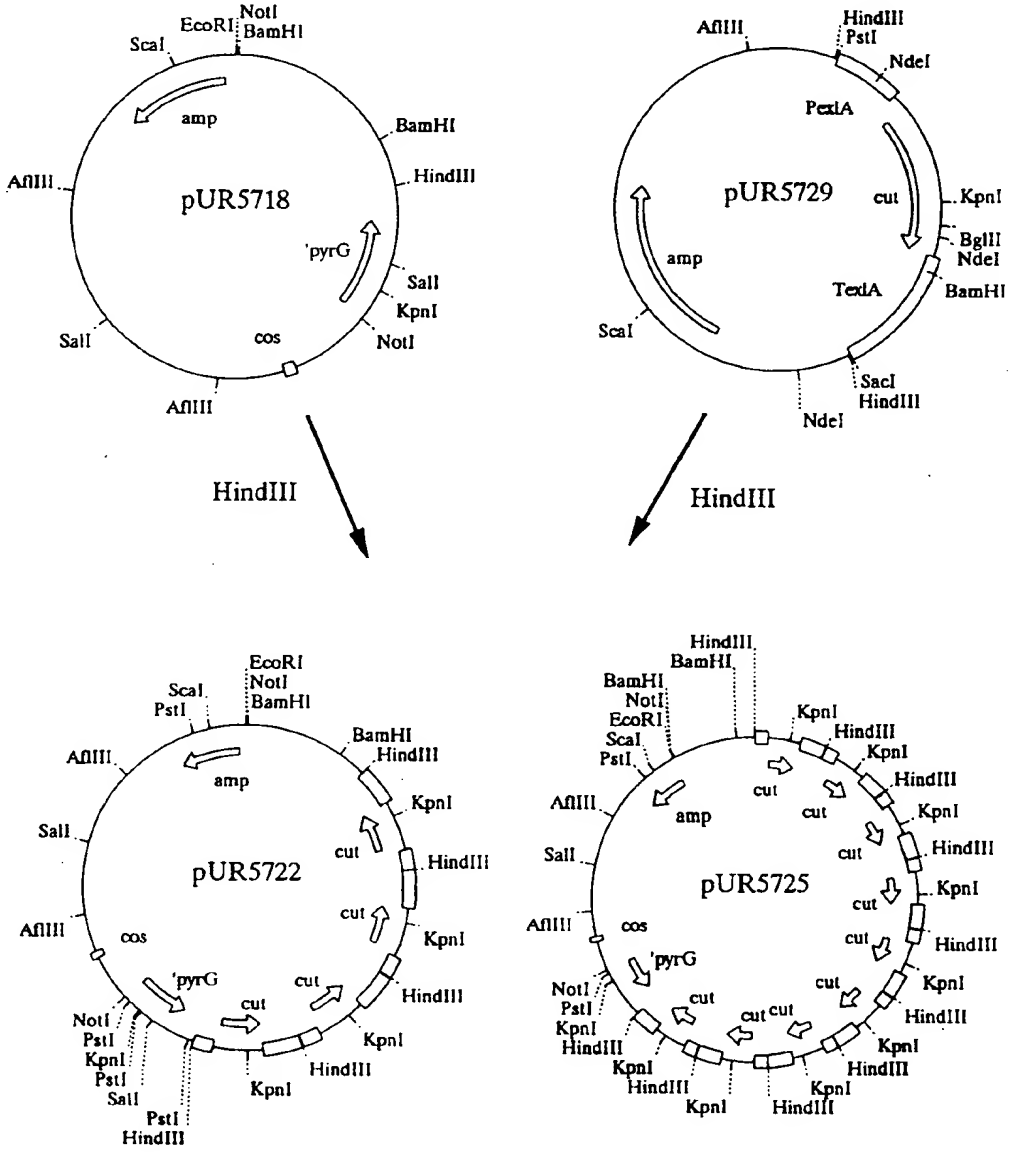

pUR5718
EcoRI
NotI
BamHI
ScaI
amp
BamHI
HindIII
AflIII
'pyrG
SalI
KpnI
cos
NotI
SalI
AflIII
pUR5729
AflIII
HindIII
PstI
NdeI
PextA
cut
KpnI
BglII
NdeI
amp
TextA
BamHI
ScaI
SacI
HindIII
NdeI
HindIII
HindIII
pUR5722
EcoRI
NotI
BamHI
ScaI
PstI
AflIII
amp
BamHI
HindIII
SalI
KpnI
cut
HindIII
AflIII
cut
KpnI
cos
'pyrG
cut
cut
HindIII
NotI
PstI
KpnI
PstI
SalI
PstI
HindIII
KpnI
HindIII
KpnI
pUR5725
HindIII
BamHI
BamHI
NotI
EcoRI
ScaI
PstI
KpnI
HindIII
KpnI
AflIII
amp
cut
cut
HindIII
KpnI
SalI
cut
HindIII
KpnI
AflIII
cos
cut
HindIII
'pyrG
cut
KpnI
NotI
PstI
KpnI
HindIII
cut
cut
cut
KpnI
HindIII
KpnI
HindIII
KpnI
HindIII

Figure 6.

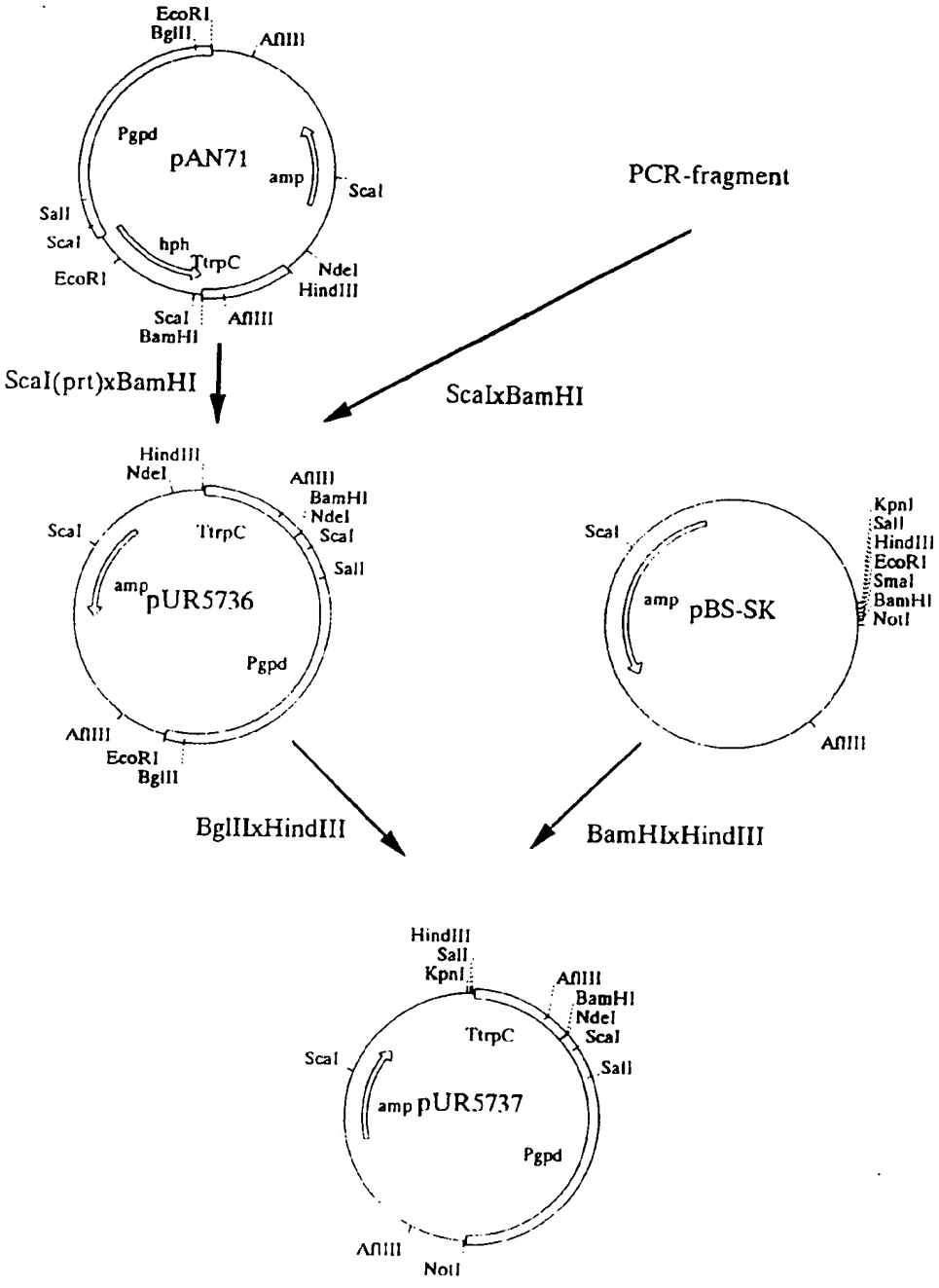

EcoRI
BglII
AflIII
Pgpd
pAN71
amp
ScaI
SalI
ScaI
hph
TtrpC
EcoRI
NdeI
HindIII
ScaI
BamHI
AflIII
ScaI(prt)xBamHI
PCR-fragment
ScaIxBamHI
HindIII
NdeI
AflIII
BamHI
NdeI
ScaI
ScaI
TtrpC
SalI
amp
pUR5736
Pgpd
AflIII
EcoRI
BglII
ScaI
amp
pBS-SK
KpnI
SalI
HindIII
EcoRI
SmaI
BamHI
NotI
AflIII
BglIIxHindIII
BamHIxHindIII
HindIII
SalI
KpnI
AflIII
BamHI
NdeI
ScaI
TtrpC
ScaI
SalI
amp
pUR5737
Pgpd
AflIII
NotI

Figure 7.

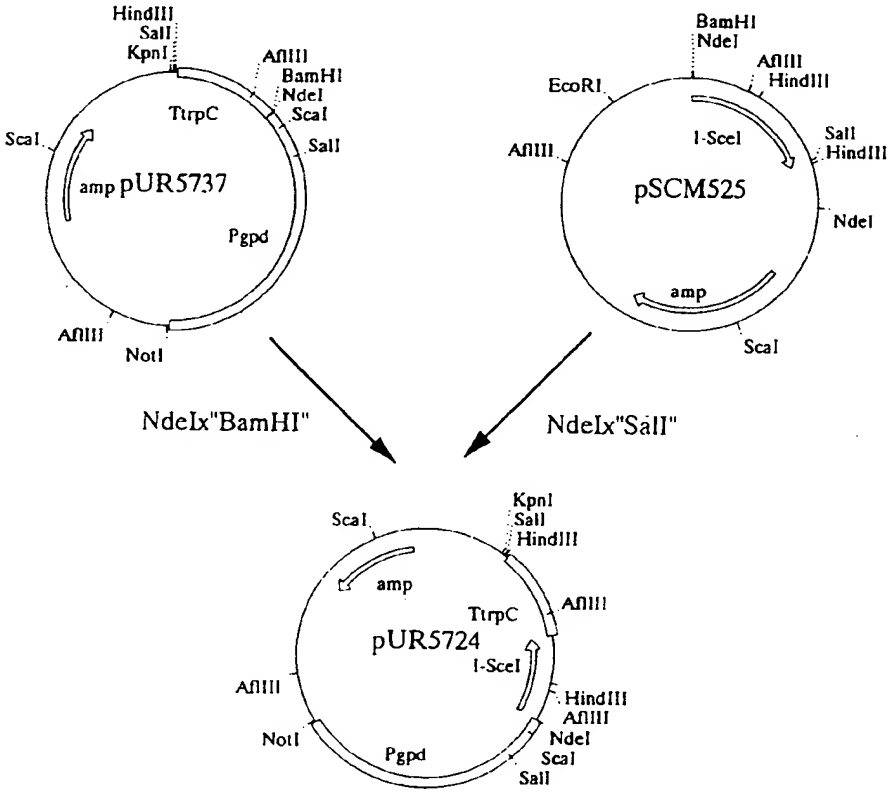

HindIII
SalI
KpnI
AflIII
BamHI
NdeI
ScaI
SalI
TtrpC
ScaI
amp pUR5737
Pgpd
AflIII
NotI
BamHI
NdeI
AflIII
HindIII
EcoRI
I-SceI
SalI
HindIII
AflIII
pSCM525
NdeI
amp
ScaI
NdeIx"BamHI"
NdeIx"SalI"
KpnI
SalI
HindIII
ScaI
amp
AflIII
TtrpC
pUR5724
I-SceI
AflIII
HindIII
AflIII
NdeI
NotI
Pgpd
ScaI
SalI

Figure 8.

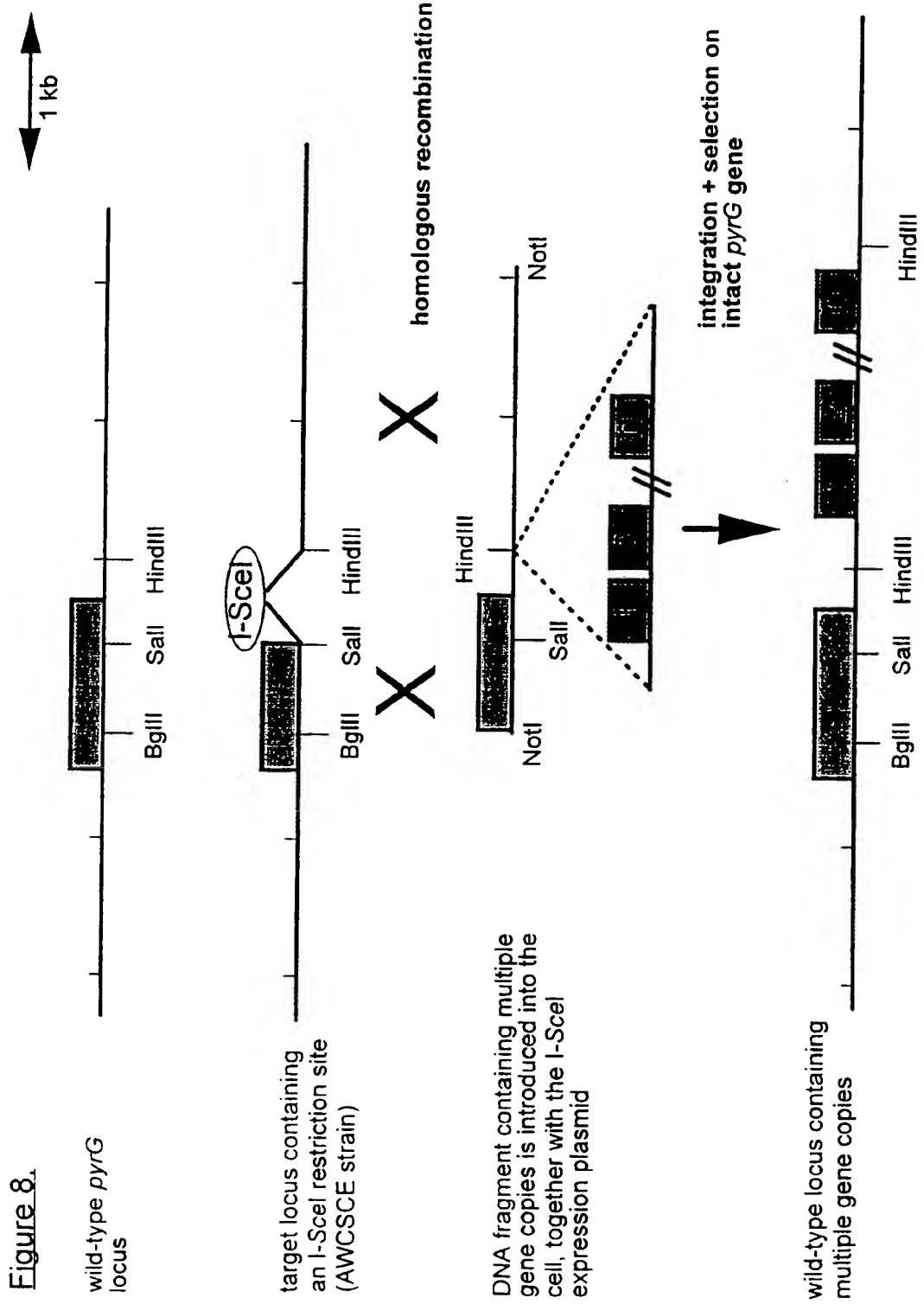

1 kb
wild-type *pyrG* locus
BglII
SalI
HindIII
I-SceI
target locus containing an I-SceI restriction site (AWCSCE strain)
homologous recombination
DNA fragment containing multiple gene copies is introduced into the cell, together with the I-SceI expression plasmid
NotI
integration + selection on intact *pyrG* gene
wild-type locus containing multiple gene copies

Figure 9

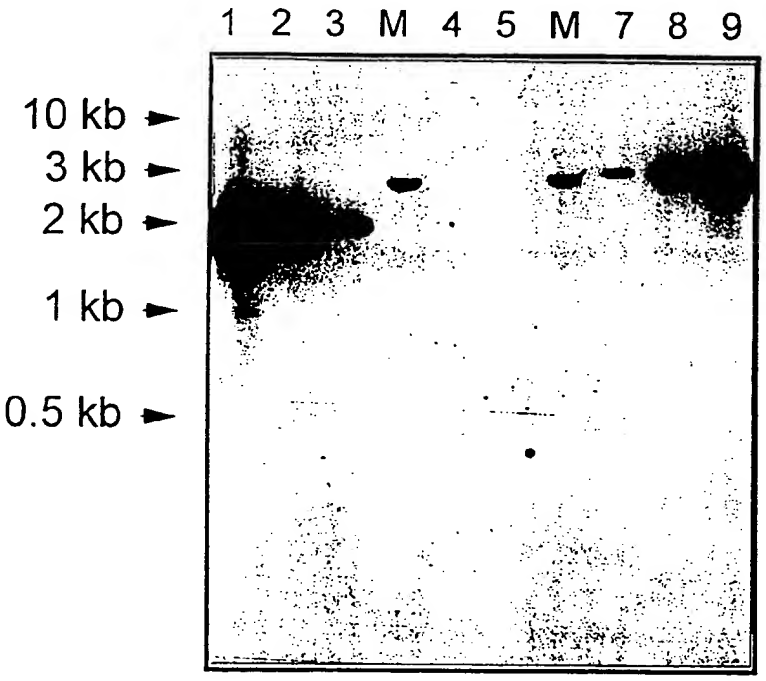


Figure 10

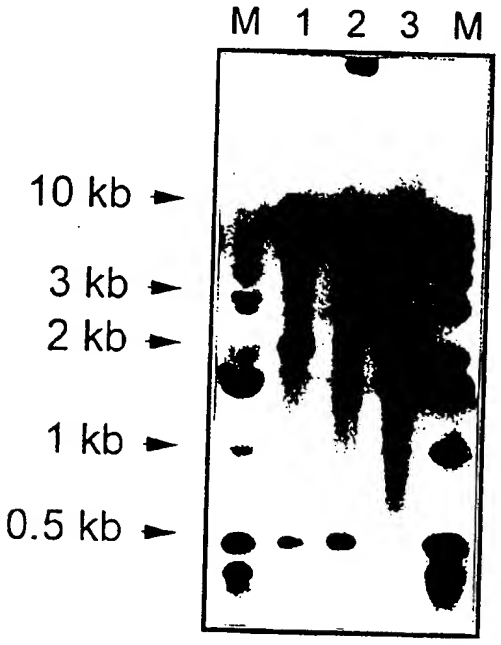


Figure 11.

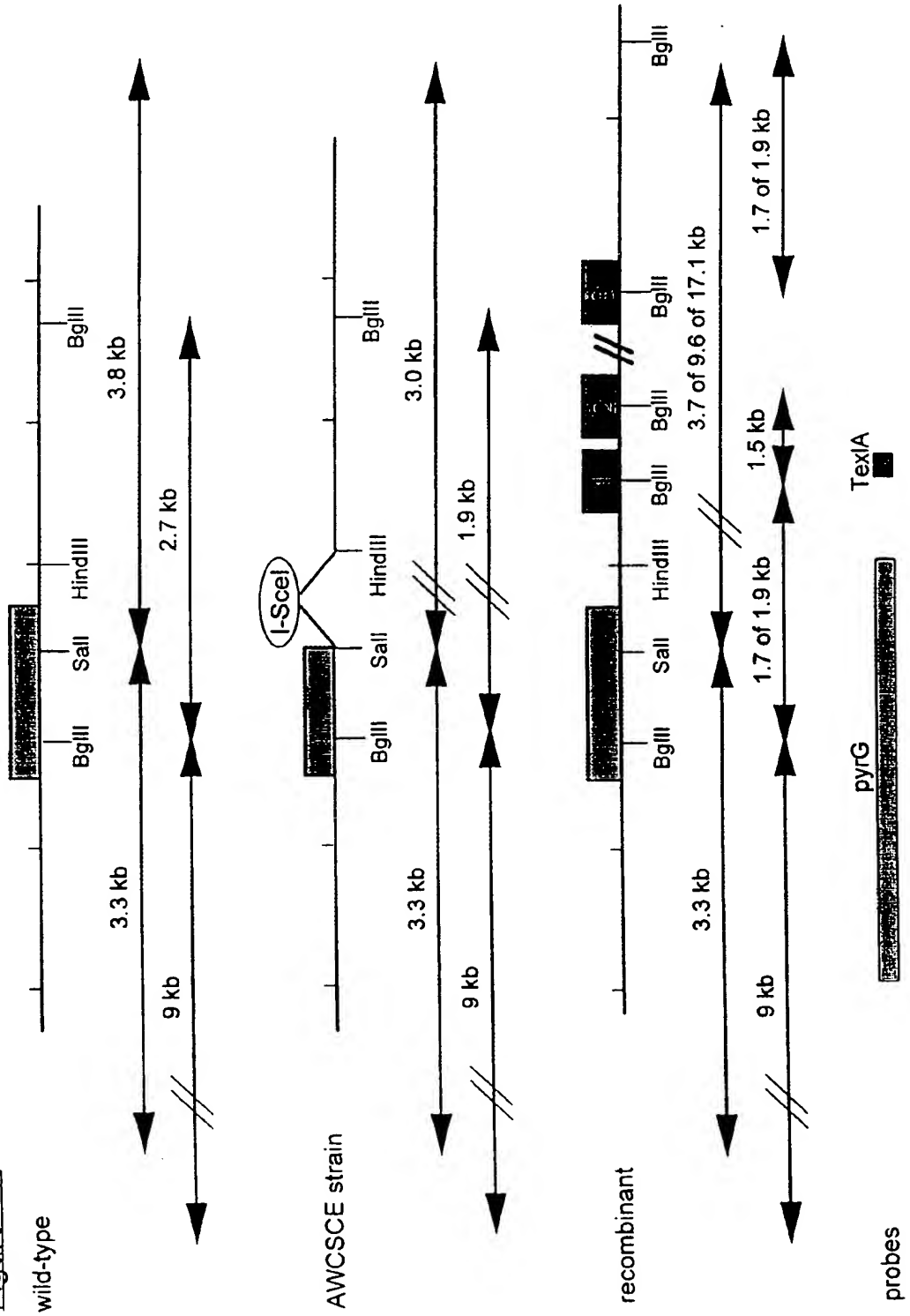
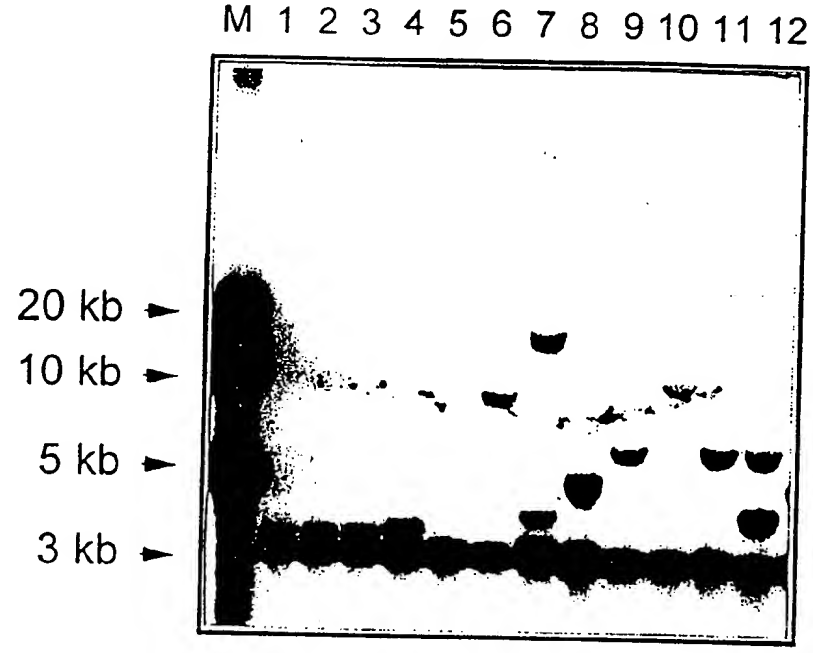
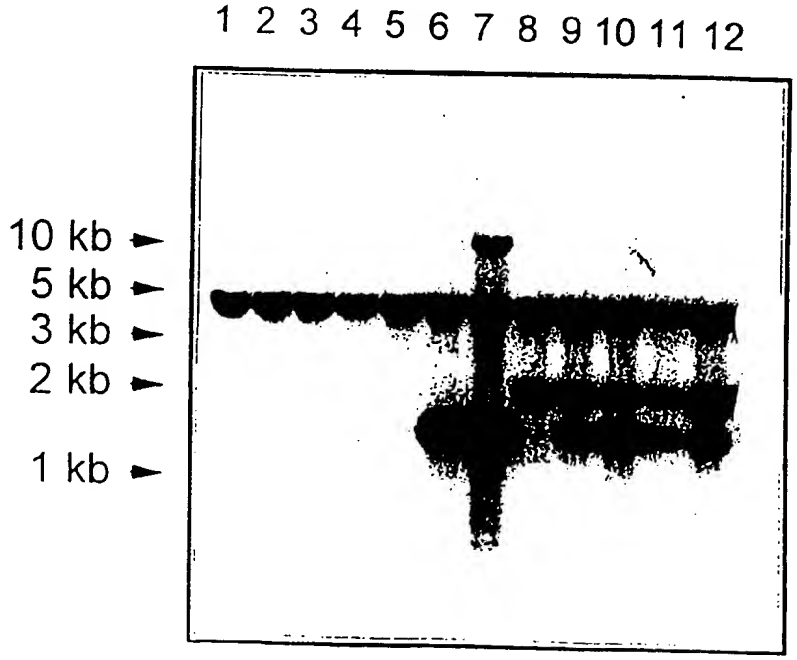

wild-type
BglII
HindIII
SalI
BglII
3.8 kb
2.7 kb
3.3 kb
9 kb
AWCSCE strain
I-SceI
BglII
HindIII
SalI
BglII
3.0 kb
1.9 kb
3.3 kb
9 kb
recombinant
BglII
HindIII
SalI
BglII
BglII
BglII
3.7 of 9.6 of 17.1 kb
1.7 of 1.9 kb
1.5 kb
1.7 of 1.9 kb
3.3 kb
9 kb
probes
pyrG
TexlA

Figure 12

A


M 1 2 3 4 5 6 7 8 9 10 11 12
20 kb
10 kb
5 kb
3 kb

B


1 2 3 4 5 6 7 8 9 10 11 12
10 kb
5 kb
3 kb
2 kb
1 kb

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 98/06519

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/80 C12N1/15 //(C12N1/15,C12R1:665,1:685,1:77,1:885)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	J.C. VERDOES ET AL.: "Glucosamylase overexpression in Aspergillus niger: molecular genetic analysis of strains containing multiple copies of the glaA gene." TRANSGENIC RESEARCH, vol. 2, 1993, pages 84-92, XP002078083 see the whole document ---	1-16
Y	T. FOWLER ET AL.: "The catR gene encoding a catalase from Aspergillus niger: primary structure and elevated expression through the increased gene copy number and use of a strong promoter" MOLECULAR MICROBIOLOGY, vol. 9, no. 5, 1993, pages 989-998, XP002041427 see the whole document ---	1-16

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 February 1999	05/03/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hix, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Ir tional Application No
PCT/EP 98/06519

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	J.C. VERDOES ET AL.: "Evaluation of molecular and genetic approaches to generate glucoamylase overproducing strains of Aspergillus niger." JOURNAL OF BIOTECHNOLOGY, vol. 36, 1994, pages 165-175, XP002093931 see the whole document	1-16
Y	R. GEISEN ET AL.: "Transformation of Penicillium nalgiovense with the amdS gene of Aspergillus nidulans." CURRENT GENETICS, vol. 15, no. 4, April 1989, pages 307-309, XP002093932 see the whole document	1-16
Y	J.M. KELLY ET AL.: "Multiple copies of the amdS gene of Aspergillus nidulands cause titration of trans-acting regulatory proteins" CURRENT GENETICS, vol. 12, 1987, pages 21-32, XP002093933 see the whole document	1-16
Y	H.J.D. BUSSINK ET AL.: "Characterization of polygalacturonase-overproducing Aspergillus niger transformants" APPLIED MICROBIOL. BIOTECH., vol. 37, no. 3, June 1992, pages 324-329, XP002093936 see the whole document	1-16
Y	WO 92 10577 A (EUROLYSINE) 25 June 1992 see the whole document	1-16
Y	EP 0 778 348 A (UNILEVER NV ;UNILEVER PLC (GB)) 11 June 1997 see the whole document	1-16
P,Y	WO 98 07873 A (KIRIN BREWERY ;KONDO KEIJI (JP); MIURA YUTAKA (JP)) 26 February 1998 see the whole document	1-16
Y	WO 91 00920 A (UNILEVER PLC ;UNILEVER NV (NL)) 24 January 1991 cited in the application see the whole document	1-16
A	EP 0 635 574 A (GIST BROCADES NV) 25 January 1995 cited in the application see the whole document	

-/--

1

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 98/06519

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	I.A. VAN GEMEREN ET AL.: "The effect of pre- and pro-sequences and multicopy integration on heterologous expression of the Fusarium solani pisi cutinase gene in Aspergillus awamori" APPL. MICROBIOL. BIOTECHNOL., vol. 45, 1996, pages 755-763, XP002093937 cited in the application see the whole document	1-16
Y	J.C. VERDOES ET AL.: "Glucoamylase overexpression in Aspergillus niger: molecular genetic analysis of strains containing multiple copies of the glaA gene." TRANSGENIC RESEARCH, vol. 2, 1993, pages 84-92, XP002093938 cited in the application see the whole document	1-16
Y	E.M. KUBICEK-PRANZ ET AL.: "Transformation of Trichoderma reesei with the cellobiohydrolase II gene as a means for obtaining strains with increased cellulase production and specific activity." JOURNAL OF BIOTECHNOLOGY, vol. 20, 1991, pages 83-94, XP002093939 cited in the application see the whole document	1-16
A	R.J. GOUKA ET AL.: "A novel strategy for the isolation of defined pyrG mutants and the development of a site-specific integration system for Aspergillus awamori." CURRENT GENETICS, vol. 27, no. 6, 1995, pages 536-540, XP002093940 cited in the application see the whole document	
Y	EP 0 249 350 A (PANLABS INC) 16 December 1987 see the whole document	1-16

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 98/06519

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9210577 A	25-06-1992	FR 2670502 A	19-06-1992
		CA 2098304 A	14-06-1992
		EP 0575338 A	29-12-1993
		JP 6504201 T	19-05-1994
EP 0778348 A	11-06-1997	EP 0481008 A	22-04-1992
		AT 160176 T	15-11-1997
		CA 2063592 A	08-01-1991
		DE 69031710 D	18-12-1997
		DE 69031710 T	05-03-1998
		DK 481008 T	04-05-1998
		WO 9100920 A	24-01-1992
		ES 2109238 T	16-01-1998
		JP 5501949 T	15-04-1993
WO 9807873 A	26-02-1998	AU 3868197 A	06-03-1998
WO 9100920 A	24-01-1991	AT 160176 T	15-11-1997
		CA 2063592 A	08-01-1991
		DE 69031710 D	18-12-1997
		DE 69031710 T	05-03-1998
		DK 481008 T	04-05-1998
		EP 0481008 A	22-04-1992
		EP 0778348 A	11-06-1997
		ES 2109238 T	16-01-1998
		JP 5501949 T	15-04-1993
EP 0635574 A	25-01-1995	AU 5302998 A	14-05-1998
		AU 690909 B	07-05-1998
		AU 6862894 A	02-02-1995
		BR 9402912 A	11-04-1995
		CA 2128589 A	24-01-1995
		CN 1098743 A	15-02-1995
		FI 943485 A	24-01-1995
		IE 940591 A	25-01-1995
		JP 7147971 A	13-06-1995
		NZ 264065 A	26-06-1998
		NZ 299139 A	26-06-1998
EP 0249350 A	16-12-1987	DK 288087 A	07-12-1987
		FI 872102 A	07-12-1987
		JP 63044890 A	25-02-1988